OM of: US-09-303-518D-569 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 30, 2002 6:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-569
Query length: 897
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000

score_list:

| Sequence | Strd | Orig | ZScore | EScore | Len | ! | Documentation | .. | | |
|---|---|---|---|---|---|---|---|---|---|---|
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74949 | + | 1580.00 | 2872.22 | 2.6e-152 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38782 | + | 1577.00 | 2866.73 | 5.3e-152 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38783 | + | 1577.00 | 2866.73 | 5.3e-152 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74950 | + | 1577.00 | 2866.73 | 5.3e-152 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60652 | + | 1573.00 | 2859.41 | 1.4e-151 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74948 | + | 1502.00 | 2729.80 | 2.3e-144 | 28 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38784 | + | 1467.50 | 2666.32 | 7.7e-141 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74947 | + | 1467.50 | 2666.32 | 7.7e-141 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38781 | + | 631.00 | 1144.41 | 1.1e-55 | 12 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60651 | + | 279.50 | 491.20 | 1.1e-19 | 30 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75338 | + | 204.50 | 354.44 | 4.7e-12 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75339 | + | 203.50 | 352.61 | 5.9e-12 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY79683 | + | 202.50 | 350.78 | 7.4e-12 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75337 | + | 197.50 | 341.62 | 2.4e-11 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98392 | + | 189.50 | 325.78 | 1.6e-10 | 32 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB88540 | + | 178.50 | 305.97 | 2.1e-09 | 31 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60654 | + | 178.50 | 305.97 | 2.1e-09 | 31 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60656 | + | 174.50 | 299.59 | 5.3e-09 | 29 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37217 | + | 172.50 | 291.18 | 1.0e-08 | 45 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60655 | + | 170.50 | 290.40 | 1.4e-08 | 34 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60653 | + | 148.50 | 251.29 | 2.4e-06 | 31 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34697 | + | 146.00 | 242.48 | 5.1e-06 | 46 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW25084 | + | 144.50 | 243.97 | 6.2e-06 | 31 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59824 | + | 131.00 | 201.80 | 0.0003 | 16 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59814 | + | 130.00 | 204.77 | 0.0003 | 10 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW85665 | + | 127.50 | 212.45 | 0.0003 | 32 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG10614 | - | 122.00 | 199.16 | 0.0014 | 43 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY51614 | + | 121.50 | 202.04 | 0.0014 | 30 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39201 | + | 121.00 | 194.60 | 0.0019 | 56 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365 | + | 121.00 | 187.53 | 0.0025 | 10 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21221 | + | 120.50 | 185.30 | 0.0029 | 12 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817 | + | 120.00 | 186.71 | 0.0030 | 99 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827 | + | 120.00 | 181.75 | 0.0036 | 15 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29221 | + | 119.00 | 193.43 | 0.0028 | 44 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21540 | + | 119.00 | 191.60 | 0.0030 | 53 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW55484 | + | 118.00 | 194.89 | 0.0032 | 32 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29184 | + | 117.50 | 188.32 | 0.0044 | 55 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB95146 | + | 117.50 | 180.21 | 0.0058 | 11 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23029 | + | 117.50 | 168.38 | 0.0087 | 36 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38613 | + | 116.50 | 192.47 | 0.0045 | 31 | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU40508 | + | 116.50 | 191.33 | 0.0047 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18664 | + | 116.50 | 190.41 | 0.0048 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813 | + | 115.00 | 177.36 | 0.0098 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826 | + | 115.00 | 172.44 | 0.0116 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB67173 | + | 113.50 | 181.51 | 0.0110 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU46556 | + | 113.00 | 188.25 | 0.0095 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB58927 | + | 113.00 | 168.28 | 0.0188 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18663 | + | 112.50 | 183.06 | 0.0124 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:AAR70649 | + | 112.50 | 182.27 | 0.0127 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB65602 | + | 112.50 | 173.47 | 0.0172 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955 | + | 112.00 | 177.97 | 0.0161 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR94957 | + | 112.00 | 171.00 | 0.0204 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04841 | + | 111.50 | 178.38 | 0.0173 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU44843 | + | 111.00 | 185.36 | 0.0148 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG08343 | + | 110.50 | 179.10 | 0.0200 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG26362 | + | 109.00 | 181.22 | 0.0241 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAG33109 | + | 109.00 | 178.13 | 0.0268 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW79128 | + | 108.00 | 178.85 | 0.0310 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38612 | + | 108.00 | 176.91 | 0.0331 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB76547 | + | 108.00 | 176.88 | 0.0332 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG91586 | + | 108.00 | 176.81 | 0.0332 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB76546 | + | 108.00 | 176.81 | 0.0332 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG14914 | + | 108.00 | 169.51 | 0.0427 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG14244 | + | 108.00 | 167.37 | 0.0459 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG08082 | + | 107.50 | 177.35 | 0.0356 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY05097 | + | 107.50 | 176.55 | 0.0366 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB87879 | + | 107.50 | 176.55 | 0.0366 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG09949 | + | 107.50 | 171.38 | 0.0436 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY50936 | + | 107.50 | 169.43 | 0.0467 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW98827 | + | 107.00 | 174.44 | 0.0428 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU62215 | + | 106.50 | 180.83 | 0.0375 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU56185 | + | 106.00 | 161.33 | 0.0797 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23142 | + | 106.00 | 161.04 | 0.0805 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE02402 | + | 105.50 | 156.56 | 0.1023 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03914 | + | 105.00 | 165.40 | 0.0823 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU22961 | + | 105.00 | 165.40 | 0.0823 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87503 | + | 105.00 | 157.19 | 0.1091 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU51580 | + | 104.50 | 163.29 | 0.0965 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731 | + | 104.50 | 162.18 | 0.1002 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU42729 | + | 104.00 | 172.90 | 0.0757 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB56814 | + | 104.00 | 171.10 | 0.0805 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU12176 | + | 104.00 | 162.95 | 0.1064 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY53044 | + | 104.00 | 161.60 | 0.1114 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21245 | + | 104.00 | 161.41 | 0.1121 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU48026 | + | 103.50 | 176.18 | 0.0737 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813 | - | 103.50 | 156.31 | 0.1455 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826 | - | 103.50 | 151.39 | 0.1722 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU42728 | + | 103.00 | 171.49 | 0.0943 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03877 | + | 103.00 | 169.70 | 0.1003 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB07854 | + | 103.00 | 165.43 | 0.1161 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG93323 | + | 103.00 | 165.43 | 0.1161 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94120 | + | 103.00 | 165.43 | 0.1161 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94427 | + | 103.00 | 165.43 | 0.1161 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB61328 | + | 103.00 | 165.43 | 0.1161 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW68408 | + | 103.00 | 165.22 | 0.1169 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB95366 | + | 103.00 | 164.02 | 0.1218 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04931 | + | 103.00 | 161.42 | 0.1331 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW87504 | + | 103.00 | 154.95 | 0.1662 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74860 | + | 102.50 | 161.57 | 0.1443 | | | | | | |
| /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAY20852 | + | 102.00 | 170.81 | 0.1147 | | | | | | |

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74949

seq_documentation_block:
ID AAY74949 standard; Protein; 298 AA.
XX
AC AAY74949;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 505 protein sequence SEQ ID NO:1372.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;

```
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53711.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 746; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1580.00              Length:    298
              Ratio:  5.302                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-303-518D-569 x AAY74949   ..

 Align seg 1/1  to: AAY74949  from: 1  to: 298

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyMetAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

     451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

     701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     851 TTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
         |||||||||||||||||||||||||||||||||||||||||||
     284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38782

seq_documentation_block:
ID   AAY38782 standard; Protein; 298 AA.
XX
AC   AAY38782;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF138.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
```

```
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12217.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 326; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1577.00              Length:    298
              Ratio:  5.292                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.664

alignment_block:
 US-09-303-518D-569 x AAY38782   ..

 Align seg 1/1  to: AAY38782  from: 1  to: 298

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         ||||||||||||||||||||||||||||||||||||||||||:::||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

     451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

     701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     851 TTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
         |||||||||||||||||||||||||||||||||||||||||||
     284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38783

seq_documentation_block:
ID   AAY38783 standard; Protein; 298 AA.
XX
AC   AAY38783;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF138.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
```

```
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12218.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 327; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1577.00              Length:    298
              Ratio:  5.292               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 99.664

alignment_block:
 US-09-303-518D-569 x AAY38783   ..

 Align seg 1/1  to: AAY38783  from: 1  to: 298

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

     451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

     701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     851 TTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
         |||||||||||||||||||||||||||||||||||||||||||
     284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74950

seq_documentation_block:
ID   AAY74950 standard; Protein; 298 AA.
XX
AC   AAY74950;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 505 protein sequence SEQ ID NO:1374.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
```

```
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53712.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 747; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1577.00              Length:    298
              Ratio:  5.292                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 99.664

alignment_block:
 US-09-303-518D-569 x AAY74950   ..

 Align seg 1/1  to: AAY74950  from: 1  to: 298

      1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

     51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

    101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

    151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
        ||||||||||||||||||||||||||||||||||||||||||||||:::||
     51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

    201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

    251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

    301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

    351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

    401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

    451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

    501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

    551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

    601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 ValProSerProGlnGluGlyGluGlyValTrpValAspPhePheGl 217

    651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

    701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

    751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

    801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

    851 TTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
        |||||||||||||||||||||||||||||||||||||||||||
    284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60652

seq_documentation_block:
ID   AAB60652 standard; Protein; 298 AA.
XX
AC   AAB60652;
XX
DT   04-MAY-2001  (first entry)
XX
DE   N. meningitidis (serogroup B) HtrB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; HtrB.
XX
OS   Neisseria meningitidis.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
```

```
DR   N-PSDB; AAF91451.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 98; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following: Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Neisseria
CC   meningitidis HtrB protein.
XX
SQ   Sequence   298 AA;


alignment_scores:
            Quality: 1573.00              Length:    298
              Ratio:  5.296                Gaps:      0
 Percent Similarity:  99.664   Percent Identity: 99.664

alignment_block:
 US-09-303-518D-569 x AAB60652   ..

 Align seg 1/1  to: AAB60652  from: 1  to: 298

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||  ||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheGlyLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyMetAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

     451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 234

     701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     851 TTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
         |||||||||||||||||||||||||||||||||||||||||||
     284 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysMetPro 298

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74948

seq_documentation_block:
ID   AAY74948 standard; Protein; 288 AA.
XX
AC   AAY74948;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 505 protein sequence SEQ ID NO:1370.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
```

```
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI   Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI   Tettelin H, Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53710.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 745; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   288 AA;


alignment_scores:
            Quality: 1502.00              Length:    286
              Ratio:  5.270                Gaps:      0
 Percent Similarity:  99.650   Percent Identity: 99.301

alignment_block:
 US-09-303-518D-569 x AAY74948   ..

 Align seg 1/1  to: AAY74948  from: 1  to: 288

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||:::||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 sHisGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheProLeuThrAlaMetTyr 150

     451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrSerIleGlnGlyValLysGlnI 184

     551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 leIleLysAlaLeuArgSerGlyGluAlaThrIleValLeuProAspHis 200

     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 ValProSerProGlnGluGlyGlyGluGlyValTrpValAspPhePheGl 217

     651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
         ||||||||||||||||||||||||||||||||  ||||||||||||||||
     217 yLysProAlaTyrThrMetThrLeuAlaAla***LeuAlaHisValLysG 234

     701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 lyValLysThrLeuPhePheCysCysGluArgLeuProGlyGlyGlnGly 250

     751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 PheAspLeuHisIleArgProValGlnGlyGluLeuAsnGlyAspLysAl 267

     801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 aHisAspAlaAlaValPheAsnArgAsnAlaGluTyrTrpIleArgArgP 284

     851 TTCCGACG 858
         ||||||||
     284 heProThr 286

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38784

seq_documentation_block:
ID   AAY38784 standard; Protein; 297 AA.
XX
AC   AAY38784;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigenic protein encoded by ORF138.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
```

```
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12219.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 328; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   297 AA;


alignment_scores:
            Quality: 1467.50              Length:    298
              Ratio:  5.078                Gaps:      1
 Percent Similarity:  96.980   Percent Identity: 93.960

alignment_block:
 US-09-303-518D-569 x AAY38784   ..

 Align seg 1/1  to: AAY38784  from: 1  to: 297

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         |||||||||||||||||||||||||||||||||||||   |||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuSerLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         |||||||||||||||||||||||||||||||||||||||||||:::|||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         |||||||   :::|||||||||||||||||||||||||||||||   |
      67 nProAspThrGlnThrValLysAlaValPheAlaGluThrAlaLysCysG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||:::|||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheLysLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         |   ||||||||||||||||||||||||||||||||||||||||||||||
     117 sGlyGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
         |||||||||||||||||||||||||||||||   |||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheHisLeuThrAlaMetTyr 150

     451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
         |||||||||||||||||||||||||||||:::|||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrGlyIleGlnGlyValLysGlnI 184

     551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
         ||||||||||||||||:::|||||||||||||||||:::|||||||||||
     184 leIleLysAlaLeuArgAlaGlyGluAlaThrIleIleLeuProAspHis 200

     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
         ||||||||||||||||||||||   ||||||||||:::|||||||||||
     201 ValProSerProGlnGluGlyGly...GlyValTrpAlaAspPhePheGl 216

     651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     216 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 233

     701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
         ||||||||||||||||||||||||||||||||||||   |||||||||
     233 lyValLysThrLeuPhePheCysCysGluArgLeuProAspGlyGlnGly 249

     751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
         |||   ||||||||||||||||||||||||||||||||||||:::|||||
     250 PheValLeuHisIleArgProValGlnGlyGluLeuAsnGlyAsnLysAl 266

     801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
         ||||||||||||||||||||||||||||:::|||||||||||||||||||
     266 aHisAspAlaAlaValPheAsnArgAsnThrGluTyrTrpIleArgArgP 283

     851 TTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
         |||||||||||||||||||||||||||||||||||||||   |||
     283 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysThrPro 297

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY74947

seq_documentation_block:
ID   AAY74947 standard; Protein; 297 AA.
XX
AC   AAY74947;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 505 protein sequence SEQ ID NO:1368.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
```

```
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53709.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 744; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   297 AA;


alignment_scores:
            Quality: 1467.50              Length:    298
              Ratio:  5.078               Gaps:        1
 Percent Similarity:  96.980   Percent Identity: 93.960

alignment_block:
 US-09-303-518D-569 x AAY74947   ..

 Align seg 1/1  to: AAY74947  from: 1  to: 297

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         |||||||||||||||||||||||||||||||||||||   |||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuSerLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||:::||
      51 LysGluAspArgAlaArgIleValAlaAsnMetArgGlnAlaGlyLeuAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         |||||||   :::|||||||||||||||||||||||||||||||   |
      67 nProAspThrGlnThrValLysAlaValPheAlaGluThrAlaLysCysG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||:::|||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheLysLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         |   ||||||||||||||||||||||||||||||||||||||||||||||||
     117 sGlyGluGlyLeuLeuPheIleThrProHisIleGlySerTyrAspLeuG 134

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
         |||||||||||||||||||||||||||||||||   ||||||||||||||
     134 lyGlyArgTyrIleSerGlnGlnLeuProPheHisLeuThrAlaMetTyr 150

     451 AAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysProProLysIleLysAlaIleAspLysIleMetGlnAlaGlyArgVa 167

     501 TCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAA 550
         ||||||||||||||||||||||||||:::|||||||||||||||||||||
     167 lArgGlyLysGlyLysThrAlaProThrGlyIleGlnGlyValLysGlnI 184

     551 TCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCAC 600
         ||||||||||||||||:::|||||||||||||||:::|||||||||||||
     184 leIleLysAlaLeuArgAlaGlyGluAlaThrIleIleLeuProAspHis 200

     601 GTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGG 650
         ||||||||||||||||||||||   ||||||||||:::||||||||||||
     201 ValProSerProGlnGluGlyGly...GlyValTrpAlaAspPhePheGl 216

     651 CAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAG 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     216 yLysProAlaTyrThrMetThrLeuAlaAlaLysLeuAlaHisValLysG 233

     701 GCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGT 750
         ||||||||||||||||||||||||||||||||||||   |||||||||
     233 lyValLysThrLeuPhePheCysCysGluArgLeuProAspGlyGlnGly 249

     751 TTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGC 800
         |||   ||||||||||||||||||||||||||||||||:::|||||
     250 PheValLeuHisIleArgProValGlnGlyGluLeuAsnGlyAsnLysAl 266

     801 CCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTT 850
         ||||||||||||||||||||||||||||:::|||||||||||||||||||
     266 aHisAspAlaAlaValPheAsnArgAsnThrGluTyrTrpIleArgArgP 283

     851 TTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAAATGCCG 894
         ||||||||||||||||||||||||||||||||||||   |||
     283 heProThrGlnTyrLeuPheMetTyrAsnArgTyrLysThrPro 297

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38781

seq_documentation_block:
ID   AAY38781 standard; Protein; 123 AA.
XX
AC   AAY38781;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF138.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
```

```
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12216.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 325; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   123 AA;


alignment_scores:
            Quality:  631.00              Length:    123
              Ratio:  5.172                Gaps:      0
 Percent Similarity:  99.187   Percent Identity: 98.374

alignment_block:
 US-09-303-518D-569 x AAY38781   ..

 Align seg 1/1  to: AAY38781  from: 1  to: 123

       1 ATGTTTCGTTTACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetPheArgLeuGlnPheArgLeuPheProProLeuArgThrAlaMetHi 17

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 sIleLeuLeuThrAlaLeuLeuLysCysLeuSerLeuLeuProLeuSerC 34

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 ysLeuHisThrLeuGlyAsnArgLeuGlyHisLeuAlaPheTyrLeuLeu 50

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         |||||||||||||||||||||||||   |||||||||||||:::||
      51 LysGluAspArgAlaArgIleValAla***MetArgGlnAlaGlyLeuAs 67

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 nProAspProLysThrValLysAlaValPheAlaGluThrAlaLysGlyG 84

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyLeuGluLeuAlaProAlaPhePheArgLysProGluAspIleGluThr 100

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetPheLysAlaValHisGlyTrpGluHisValGlnGlnAlaLeuAspLy 117

     351 ACACGAAGGGCTGCTATTC 369
         |||||||||||||||||||
     117 sHisGluGlyLeuLeuPhe 123

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60651

seq_documentation_block:
ID   AAB60651 standard; Protein; 308 AA.
XX
AC   AAB60651;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Moraxella catarrhalis HtrB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; HtrB.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91450.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 97; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Moraxella catarrhalis
CC   HtrB protein.
XX
SQ   Sequence   308 AA;


alignment_scores:
            Quality:  279.50              Length:    285
              Ratio:  1.644                Gaps:      8
 Percent Similarity:  59.649   Percent Identity: 30.175

alignment_block:
 US-09-303-518D-569 x AAB60651   ..

 Align seg 1/1  to: AAB60651  from: 1  to: 308

      64 GCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCT 113
         ::::::|||      |||      :::||||||||||   ||||||   ||
```

```
     1 SerValLeuGlyPheLeuArgTyrValProLeuSerValLeuHisGlyLe 17

   114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
       |:::          :::::::::::::::::          :::
    17 uAlaAlaCysAlaSerTyrIleSerTyrHisCysArgLeuSerIleTyrA 34

   164 CGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAAT...CCCGAC... 207
         |||   ||||||:::      :::      :::   ||||||
    34 rgSerIleGlnAlaAsnLeuIleLeuValHisProLysMetProAspAla 50

   208 ......CCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGG 251
             |||   :::|||   ::::::   :::        :::::
    51 GlnArgGlnLysLeuAlaLysGlnIleLeuLysAsnGlnLeuIleSerAl 67

   252 TTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAA 301
       :::::::         ::::::         |||:::         ::::
    67 aValAspSerLeuLysThrTrpAlaMetProProLysTrpSerIleAlaG 84

   302 TGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAA 351
       :::::|||:::||||||      :::   :::   :::||||||   :::
    84 lnIleLysThrValHisHisGluAspIleLeuIleLysAlaLeuAlaAsn 100

   352 CACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
          :::|||:::|||   |||:::||||||||||||:::::::::::::
   101 ProSerGlyMetLeuAlaIleValProHisIleGlyThrTrpGluMetMe 117

   402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACA 451
        :::   :::::::::      :::      |||   |||   |||||||
   117 tAsnAlaTrpLeuAsn...ThrPheGlySerPro...ThrIleMetTyrL 132

   452 AACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTT 501
       |||||   |||      |||:::|||:::::::::      ||||||
   132 ysProIleLysAsnAlaAlaValAspArgPheValLeuGlnGlyArgGlu 148

   502 CGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAAT 551
       |||   :::::::::   :::|||||||:::   :::|||||||||   ||
   149 ArgLeuAsnAlaSerLeuValProThrAspAlaSerGlyValLysAlaIl 165

   552 CATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCACG 601
       |:::|||:::|||:::::|||      :::|||:::|||||||||||||
   165 ePheLysThrLeuLysAlaGlyGlyPheSerIleIleLeuProAspHisV 182

   602 TCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGC 651
       |||||:::|||   :::|||||||||         :::   |||||||||
   182 alProAspPro...SerGlyGlyGlu......IleAlaProPhePheGly 195

   652 AAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGG 701
             :::   |||   |||||||||:::||||||||         ||
   196 IleLysThrLeuThrSerThrLeuAlaSerLysLeuAlaAlaLysThrGl 212

   702 CGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTT 751
       |       :::   :::   |||   |||         |||:::||||
   212 yCysAlaLeuValGlyLeuSerCysIleArgArgGluAspGlyAspGlyP 229

   752 TCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGCC 801
       ||::::::                  |||||||||   ::::::
   229 heGluIlePheCysTyr............GluLeuAsnAspGluGlnLeu 241

   802 CATGAT.................GCCGCCGTGTTCAACCGCAATGCCGA 833
       ::::::                 ::::::::::::|||         ||
   242 TyrSerLysAsnThrLysIleAlaThrThrAlaLeuAsnGlyAlaMetGl 258

   834 ATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCT 883
       |      |||              :::|||:::::   |||:::|||:
   258 uGlnMetIleTyrProHisPheLeuHisTyrMetTrpSerTyrArgArgP 275

   884 ACAAA 888
       ::|||
   275 heLys 276

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75338

seq_documentation_block:

ID   AAY75338 standard; Protein; 293 AA.
XX
AC   AAY75338;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 663 protein sequence SEQ ID NO:2150.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54100.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1054; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   293 AA;


alignment_scores:
            Quality:  204.50              Length:    286
              Ratio:  1.225                 Gaps:     11
 Percent Similarity:  58.392   Percent Identity: 26.923

alignment_block:
 US-09-303-518D-569 x AAY75338   ..

 Align seg 1/1  to: AAY75338  from: 1  to: 293

      67 CTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGG 116
         :::|||      |||::::::||||||::::::   ||||||   :::::
```

```
 11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysIleAl 27

117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
    ::::      |||   ||||||   |||||||||   :::   |||   |
 27 aAspLeuThrGlyLeuLeuAla...TyrLeuLeuValLysProArgArgA 43

167 GCATC............GTCGCCAATATGCGTCAGGCAGGCATGAATCCC 204
    |||||           :::|||:::      :::         :::
 43 rgIleGlyGluIleAsnLeuAlaLysCysPheSerGluTrpSerGluGlu 59

205 GACCCCAAAACGGTC...AAAGCCGTTTTTGCGGAAACGGCAAAAGGCGG 251
          |||||||||   |||      |||   :::   ||||||
 60 LysArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMe 76

252 TTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAA 301
     ||||||   :::      ::::::           ::::::::::
 76 tLeuGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerL 93

302 TGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAA 351
    ::   :::      ::::::   ::::::::::::::::||||||
 93 euValArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAla 108

352 CACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
       |||   ::::::::::::   |||||||:::   ::::::::::::::
109 GlyGluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAl 125

402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACA 451
    :         ::::::|||::::::       ||||||   :::|||||||:
125 aValTyrAlaLeuAsnGlnAspIle......ProLeuIleSerMetTyrS 140

452 AACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTT 501
    ::      |||   |||   :::|||:::   :::      ||||||
140 erHisGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsn 156

502 CGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAAT 551
    |||      :::                          :::|||::::::   ::
157 ArgTyrHisAsnValPheLeuIleGlyArgThrGluGlyLeuArgAlaLe 173

552 CATCAAAGCCCTGCGTTCGGGCGAAGCAACC...ATCGTCCTGCCCGACC 598
    ::::|||   :::|||:::::::::|||      :::   |||||||||
173 uValLysGlnPheArgLysSerSerAlaProPheLeuTyrLeuProAsp. 189

599 ACGTCCCCTCCCCTCAAGAAGGCGGG......GAAGGCGTATGGGTGGAT 642
                ||||::   |||      ::::::|||:::||||||
190 ..............GlnAspPheGlyArgAsnAspSerValPheValAsp 201

643 TTCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACA 692
    |||||||||      :::   |||:::|||      ::::::::|||
202 PhePheGlyIleGlnThrAlaThrIleThrGlyLeuSerArgIleAlaAl 218

693 CGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCT.... 738
     :::   ::::::|||::::::                 :::|||
218 aLeuAlaAsnAlaLysValIle.............ProAlaIleProValA 231

739 .....GGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAA 783
         :::   :::         ||||||:::   |||:::      :::
231 rgGluAlaAspAsnThrValThrLeuHisPheTyrProAlaTrpLysSer 247

784 TTGAACGGCGACAAAGCCCAT...GATGCCGCCGTGTTCAACCGCAATGC 830
    :::   |||:::   |||      ||||||      :::||||||
248 PheProGlyGluAspAlaLysAlaAspAlaGlnArgMetAsnArgPheIl 264

831 CGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACC 880
     |||      :::|||:::   |||   |||||||:::::::::::::::|
264 eGluAspArgValArgGluHisProGluGlnTyrPheTrpLeuHisLysA 281

881 GCTACAAA 888
    ||:::|||
281 rgPheLys 283
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75339

seq_documentation_block:

```
ID   AAY75339 standard; Protein; 293 AA.
XX
AC   AAY75339;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 663 protein sequence SEQ ID NO:2152.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54101.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1055; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   293 AA;
```

```
alignment_scores:
            Quality:  203.50              Length:    285
              Ratio:  1.219                 Gaps:      9
 Percent Similarity:  58.596   Percent Identity: 25.965

alignment_block:
 US-09-303-518D-569 x AAY75339   ..

 Align seg 1/1  to: AAY75339  from: 1  to: 293

      67 CTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGG 116
         :::|||      |||::::::|||||::::::   ||||||   |||::
```

```
 11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysLeuAl 27

117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
    ::::      |||   ||||||:::   |||:::|||      |||
 27 aAspLeuThrGlyLeuLeuAlaTyrLeuLeuValLysProArgArgArgI 44

167 GCATCGTCGCCAATATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACG 216
              |||:::   ::::::      :::   |||   |||
 44 leGlyGluIleAsnLeuAlaLysCysPheProGluTrpAspGlyLysLys 60

217 GTCAAAGCCGTT...........TTTGCGGAAACGGCAAAAGGCGGTTT 254
       |||:::|||          |||   :::   ||||||      ||
 61 ArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMetLe 77

255 GGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGT 304
    ||||   :::      ::::::                :::::::::::
 77 uGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerLeuV 94

305 TCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACAC 354
      :::      ::::::   ::::::::::::::||||||
 94 alArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAlaGly 109

355 GAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
    |||   ::::::::::::   ||||||:::   ::::::::::::::
110 GluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAlaVa 126

405 ACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAAC 454
          :::::::|||:::::      ||||||   :::||||||:::
126 lTyrAlaLeuAsnGlnAspVal......ProLeuIleSerMetTyrSerH 141

455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGC 504
         |||   |||   :::|||:::   :::      ||||||   |||
141 isGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsnArg 157

505 GGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCAT 554
          :::                      :::|||:::::   :::::
158 TyrHisAsnValPheLeuIleGlyArgThrGluGlyLeuArgAlaLeuVa 174

555 CAAAGCCCTGCGTTCGGGCGAAGCAACC...ATCGTCCTGCCCGACCACG 601
    :|||   :::|||::::::::::|||      :::   |||||||||
174 lLysGlnPheArgLysSerSerAlaProPheLeuTyrLeuProAsp.... 189

602 TCCCCTCCCCTCAAGAAGGCGGG......GAAGGCGTATGGGTGGATTTC 645
              |||:::   |||      ::::::|||:::|||||||||
190 ...........GlnAspPheGlyArgAsnAspSerValPheValAspPhe 202

646 TTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGT 695
    ||||||      :::   |||:::|||      :::::::::|||   ::
203 PheGlyIleArgThrAlaThrIleThrGlyLeuSerArgIleAlaAlaLe 219

696 CAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCT..... 738
    :   ::::::|||::::::                  :::|||
219 uAlaAsnAlaLysValIle............ProAlaIleProValArgG 232

739 ..GGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTG 786
      :::   :::         ||||||:::   |||:::      ::::::
232 luAlaAspAsnThrValThrLeuHisPheTyrProAlaTrpGluSerPhe 248

787 AACGGCGACAAAGCCCAT...GATGCCGCCGTGTTCAACCGCAATGCCGA 833
       ::::::   |||:::   ||||||      :::||||||     ||
249 ProSerGluAspAlaGlnAlaAspAlaGlnArgMetAsnArgPheIleGl 265

834 ATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCT 883
    |      :::|||:::   |||   ||||||:::::::::::::::|||:
265 uGluArgValArgGluHisProGluGlnTyrPheTrpLeuHisLysArgP 282

884 ACAAA 888
    ::|||
282 heLys 283
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY79683

seq_documentation_block:

```
ID   AAY79683 standard; Protein; 293 AA.
XX
AC   AAY79683;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Neisseria meningitidis HtrB1 acyltransferase.
XX
KW   Lipopolysaccharide; vaccine; adjuvant; htrB1 gene;
KW   acyltransferase; toxicity; attenuation.
XX
OS   Neisseria meningitidis.
XX
PN   WO200026384-A1.
XX
PD   11-MAY-2000.
XX
PF   03-NOV-1998;   98WO-NL00633.
XX
PR   03-NOV-1998;   98WO-NL00633.
XX
PA   (NEWE-) NEDERLANDEN MIN WELZIJN.
XX
PI   Van Der Ley PA,  Hamstra HJ,  Steeghs LJJM;
XX
DR   WPI; 2000-422514/36.
DR   N-PSDB; AAA27756.
XX
PT   New recombinant lipopolysaccharide, useful as low-toxicity adjuvant for
PT   vaccines, has altered pattern of acylation and/or phosphate residues
PT   attached to glucosamine -
XX
PS   Disclosure; Fig 2B; 40pp; English.
XX
CC   The present sequence is that of an acyltransferase encoded by the
CC   htrB1 gene (see AAA27756) of Neisseria meningitidis.  The
CC   acyltransferase is involved in the secondary acylation of
CC   lipopolysaccharide (LPS).  Mutations in the htrB1 gene provide
CC   an LPS product that is less toxic than native LPS, but has higher
CC   adjuvant activity.  The invention is directed at novel less toxic
CC   forms of LPS that are obtained through genetically modified
CC   Gram-negative bacteria.  The novel LPS has fewer secondary acyl
CC   chains per molecule of LPS than the native LPS, the secondary
CC   acyl chains being bound to primary acyl chains, and the primary
CC   acyl chains being bound to the glucosamine of the LPS molecule.
CC   Recombinant LPS is produced by cultivation of a Gram-negative
CC   bacterium, such as N. meningitidis, having a mutation in a gene
CC   encoding a protein involved in lipid A biosynthesis, particularly
CC   at the level of secondary acyl addition, especially the htrB1 gene.
CC   It is used as an adjuvant in vaccines used to stimulate an immune
CC   response against Gram-negative bacteria, particularly for
CC   controlling infections caused by organisms from which the LPS is
CC   derived, or by other organisms.  The acylation pattern of the
CC   recombinant LPS is homogeneous, which facilitates standardization.
XX
SQ   Sequence   293 AA;
```

```
alignment_scores:
            Quality:  202.50              Length:    285
              Ratio:  1.227                 Gaps:     10
 Percent Similarity:  57.895   Percent Identity: 25.965

alignment_block:
 US-09-303-518D-569 x AAY79683   ..

 Align seg 1/1  to: AAY79683  from: 1  to: 293

      67 CTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGG 116
         :::|||      |||::::::||||||::::::   ||||||   :::::
      11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysIleAl 27

     117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
```

```
      27 aAspLeuThrGlyLeuLeuAlaTyrLeuLeuValLysProArgArgArgT 44

     167 GCATCGTCGCCAATATGCGTCAGGCA.........GGCATGAATCCCGAC 207
      44 hrGlyGluIleAsnLeuAlaLysCysPheSerGluTrpSerGluGluLy# 60

     208 CCCAAAACGGTC...AAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTTT 254
      61 ArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMetLe 77

     255 GGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGT 304
      77 uGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerLeuV 94

     305 TCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACAC 354
      94 alArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAlaGly 109

     355 GAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
     110 GluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAlaVa 126

     405 ACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAAC 454
     126 lTyrAlaLeuAsnGlnAspIle......ProLeuIleSerMetTyrSerH 141

     455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGC 504
     141 isGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArgAsnArg 157

     505 GGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCAT 554
     158 TyrHisAsnValPheLeuIleGlyArgThrGluGlyLeuArgAlaLeuVa 174

     555 CAAAGCCCTGCGTTCGGGCGAAGCAACC...ATCGTCCTGCCCGACCACG 601
     174 lLysGlnPheArgLysSerSerAlaProPheLeuTyrLeuProAsp.... 189

     602 TCCCCTCCCCTCAAGAAGGCGGG......GAAGGCGTATGGGTGGATTTC 645
     190 ...........GlnAspPheGlyArgAsnAspSerValPheValAspPhe 202

     646 TTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGT 695
     203 PheGlyIleGlnThrAlaThrIleThrGlyLeuSerArgIleAlaAlaLe 219

     696 CAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCT.....  738
     219 uAlaAsnAlaLysValIle.....ProAlaIleProValArgG 232

     739 ..GGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTG 786
     232 luAlaAspAsnThrValThrLeuHisPheTyrProAlaTrpLysSerPhe 248

     787 AACGGCGACAAAGCCCAT...GATGCCGCCGTGTTCAACCGCAATGCCGA 833
     249 ProGlyGluAspAlaLysAlaAspAlaGlnArgMetAsnArgPheIleGl 265

     834 ATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCT 883
     265 uAspArgValArgGluHisProGluGlnTyrPheTrpLeuHisLysArgP 282

     884 ACAAA 888
     282 heLys 283

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75337

seq_documentation_block:
ID   AAY75337 standard; Protein; 293 AA.
XX
AC   AAY75337;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 663 protein sequence SEQ ID NO:2148.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54099.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 1053; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   293 AA;


alignment_scores:
            Quality:  197.50              Length:    297
              Ratio:  1.197                 Gaps:     11
 Percent Similarity:  55.556   Percent Identity: 25.253

alignment_block:
 US-09-303-518D-569 x AAY75337   ..

 Align seg 1/1  to: AAY75337  from: 1  to: 293

      67 CTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGG 116
      11 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysIleAl 27

     117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
```

```
     ::::   :::|||   ||||||:::   |||:::|||      |||
 27 aGlyLeuIleGlySerLeuAlaTyrLeuLeuValLysProArgArgArg. 43

167 GCATCGTCGCCAATATGCGTCAGGCAGGCATGAATCCCGAC..... 207
          ::::::::::::::   |||         |||:::
 44 .....IleGlyGluIleAsnLeuAlaLysCysPheProGluTrpAspGlu 58

208 ......CCCAAAACGGTC...AAAGCCGTTTTTGCGGAAACGGCAAAAGG 248
           |||||||||   |||      |||   :::   ||||||
 59 GluLysArgLysThrValLeuLysGlnHisPheLysHisMetAlaLysLe 75

249 CGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAA 298
       ||||||   :::      ::::::   :::   :::   ::::::
 75 uMetLeuGluTyrGlyLeuTyrTrpTyrAlaSerAlaLysCysLeuLysS 92

299 CAATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGAC 348
     :::::   :::      ::::::   ::::::::::::::||||||
 92 erLeuValArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAla 107

349 AAACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTT 398
         |||   ::::::::::::   |||||||:::   ::::::::::
108 AlaGlyGluLysValIleIleLeuTyrProHisPheThrAlaPheGluMe 124

399 GGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGT 448
     ::::         ::::::|||:::::::      ||||||   :::||||
124 tAlaValTyrAlaLeuAsnGlnAspVal......ProLeuIleSerMetT 139

449 ACAAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGG 498
     ||:::      |||   |||   :::|||:::   :::      ||||||
139 yrSerHisGlnLysAsnLysIleLeuAspGluGlnIleLeuLysGlyArg 155

499 GTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACA 548
        |||      :::                    :::|||:::::
156 AsnArgTyrHisAsnValPheLeuIleGlyArgThrGluGlyLeuArgAl 172

549 AATCATCAAAGCCCTGCGTTCGGGCGAAGCAACC...ATCGTCCTGCCCG 595
      :::::::|||   :::|||::::::::::|||      :::   |||||||
172 aLeuValLysGlnPheArgLysSerSerAlaProPheLeuTyrLeuProA 189

596 ACCACGTCCCCTCCCCTCAAGAAGGCGGG......GAAGGCGTATGGGTG 639
     ||               |||:::   |||      ::::::|||:::|||
189 sp..............GlnAspPheGlyArgAsnAsnSerValPheVal 200

640 GATTTCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGC 689
     ||||||||||||      :::   |||:::|||      ::::::::::||
201 AspPhePheGlyIleGlnThrAlaThrIleThrGlyLeuSerArgIleAl 217

690 ACACGTCAAAGGCGTGAAA.............................. 708
     |   :::   ::::::|||
217 aAlaLeuAlaAsnAlaLysValIleProAlaIleProValArgGluAlaA 234

709 ...........ACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAA 747
              ||||||   |||              |||:::   :::
234 spAsnThrValThrLeuGlnPhe...........TyrProAlaTrpLys 246

748 GGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAA 797
     :::|||                                :::   |||
247 SerPhePro..........................SerGluAspAl 253

798 AGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCC 847
          ||||||      :::||||||      |||      :::|||:
253 aGlnAlaAspAlaGlnArgMetAsnArgPheIleGluGluArgValArgG 270

848 GTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAA 888
     ::   |||   ||||||:::::::::::::::|||:::|||
270 luHisProGluGlnTyrPheTrpLeuHisLysArgPheLys 283
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98392

seq_documentation_block:

ID AAG98392 standard; Protein; 328 AA.
XX
AC AAG98392;
XX
DT 21-SEP-2001 (first entry)
XX
DE Escherichia coli protein sequence SEQ ID NO:440.
XX
KW Escherichia coli; identification; proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW bacterial growth inhibition.
XX
OS Escherichia coli.
XX
PN WO200148209-A2.
XX
PD 05-JUL-2001.
XX
PF 19-DEC-2000; 2000WO-US34419.
XX
PR 23-DEC-1999; 99US-0173005.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen KL, Zyskind JW;
XX
DR WPI; 2001-457376/49.
DR N-PSDB; AAH81448.
XX
PT Novel nucleic acids encoding proteins required for Escherichia coli
PT proliferation, useful for screening for antimicrobial agents -
XX
PS Claim 19; Page 559; 596pp; English.
XX
CC The present invention describes a purified or isolated nucleic acid
CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (I) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools.
CC For example, nucleic acid probes complementary to proliferation-required
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.
XX
SQ Sequence 328 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 189.50 | Length: | 294 |
| Ratio: | 1.199 | Gaps: | 12 |
| Percent Similarity: | 53.741 | Percent Identity: | 26.190 |

alignment_block:

US-09-303-518D-569 x AAG98392 ..

Align seg 1/1 to: AAG98392 from: 1 to: 328

```
     88 CTGCCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGC 137
        ||||||         |||      ||||||:::|||:::|||   :::||
     54 LeuProTyrProValLeuCysPheLeuGlyThrArgIleGlyAlaMetAl 70

    138 GTTTTACCTTTTAAAG.............................. 153
         |      :::|||||||
     70 aArgProPheLeuLysArgArgGluSerIleAlaArgLysAsnLeuGluL 87
```

```
 154 .....GAAGACCGCGCGCGCATCGTCGCC... 177
                        |||:::|||   :::::::::|||
  87 euCysPheProGlnHisSerAlaGluGluArgGluLysMetIleAlaGlu 103

 178 AATATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGCCGT 227
     |||:::|||:::   ||||||
 104 AsnPheArgSerLeuGlyMet..... 110

 228 TTTTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCA 277
     |                 :::   :::|||   :::   |||:::|||
 111 ..................AlaLeuValGluThrGlyMetAlaTrpPheT 121

 278 GAAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTACACGGCTGGGAA 327
           :::   ::::::      |||      |||:::|||   :::
 121 rpProAspSerArgValArgLysTrpPheAsp...ValGluGlyLeuAsp 136

 328 CATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTATTCATCACGCC 377
     ::::::::::::|||     :::::::::|||::::::   :::
 137 AsnLeuLysArgAlaGlnMetGlnAsnArgGlyValMetValValGlyVa 153

 378 GCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTC 427
      |||:::   |||   :::|||||||||||||   ::::::      |||
 153 lHisPheMetSerLeuGluLeuGlyGlyArgValMetGly......LeuC 168

 428 CGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCCATAGAC 477
          |||:::   |||   |||:::|||   :::   :::   ::::::
 168 ysGlnProMetMetAlaThrTyrArgProHisAsnAsnGlnLeuMetGlu 184

 478 AAAATCATGCAGGCGGGCAGGGTTCGC...GGCAAAGGAAAAACCGCGCC 524
        ::::::      ||||||:::|||   :::|||:::      :::
 185 TrpValGlnThrArgGlyArgMetArgSerAsnLysAlaMetIleGlyAr 201

 525 TACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCG 574
      :::::::::::|||         |||:::   ||||||::::::||||
 201 gAsnAsnLeuArgGly.........IleValGlyAlaLeuLysLysGlyG 215

 575 AAGCAACCATCGTCCTGCCCGACCACGTCCCCTCCCCTCAAGAAGGCGGG 624
     |||||:::         ||||||                 |||:::   |||
 215 luAlaValTrpPheAlaProAsp..............GlnAspTyrGly 226

 625 GAA......GGCGTATGGGTGGATTTCTTCGGC...AAACCTGCCTATAC 665
     :::      :::   ::::::   ||||||:::   :::   :::   ||
 227 ArgLysGlySerSerPheAlaProPhePheAlaValGluAsnValAlaTh 243

 666 CATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGT 715
     |   :::   :::      |||:::::::::::::|||:::      |||
 243 rThrAsnGlyThrTyrValLeuSerArgLeuSerGlyAlaAlaMetLeuT 260

 716 TTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATC 765
                |||            :::|||:::   |||   |||
 260 hrValThrMetValArgLysAlaAspTyrSerGlyTyrArgLeuPheIle 276

 766 CGCCCCGTCCAAGGGGAATTGAACGGC.........GACAAAGCCCATGA 806
        |||         |||:::::::|||         |||:::   :::
 277 ThrPro.........GluMetGluGlyTyrProThrAspGluAsnGlnAl 290

 807 TGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTTTTCCGA 856
      ||||||   :::|||:::      |||      |||   |||   |||
 290 aAlaAlaTyrMetAsnLysIleIleGluLysGluIleMetArgAlaProG 307

 857 CGCAGTATCTGTTTATGTACAACCGCTACAAA 888
       |||||||||:::::::::::::|||:::|||
 307 luGlnTyrLeuTrpIleHisArgArgPheLys 317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB88540

seq_documentation_block:

ID AAB88540 standard; Protein; 318 AA.
XX
AC AAB88540;
XX
DT 04-JUN-2001 (first entry)
XX
DE Haemophilus influenzae essential bacterial protein SEQ ID NO:98.
XX
KW Haemophilus influenzae; essential bacterial gene; identification;
KW otitis media; meningitis; upper respiratory tract infection;
KW infection; antimicrobial.
XX
OS Haemophilus influenzae.
XX
PN WO200111033-A2.
XX
PD 15-FEB-2001.
XX
PF 03-AUG-2000; 2000WO-US21176.
XX
PR 04-AUG-1999; 99US-0368382.
XX
PA (ABBO ) ABBOTT LAB.
XX
PI Chovan LE, Hessler PE, Reich KA;
XX
DR WPI; 2001-147511/15.
DR N-PSDB; AAF94393.
XX
PT Essential bacterial genes from Haemophilus influenzae and methods for
PT identifying 'essential' genes that may be potential therapeutic targets
PT -
XX
PS Claim 9; Page 149; 185pp; English.
XX
CC AAF94345 to AAF94409 represent essential bacterial genes from
CC Haemophilus influenzae, which encode the proteins given in AAB88492 to
CC AAB88556. The present invention also describes methods for identifying
CC essential bacterial genes (i.e. those essential to the survival of a
CC bacterium) using a transposition system. The methods are used to
CC identify essential genes from bacteria, especially H. influenzae (which
CC causes otitis media, meningitis and upper respiratory tract infections)
CC which may be used as targets for potential antimicrobial agents.
CC AAF94410 to AAF94416 represent PCR primers used in the exemplification
CC of the present invention.
XX
SQ Sequence 318 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 178.50 | Length: | 275 |
| Ratio: | 1.137 | Gaps: | 9 |
| Percent Similarity: | 57.091 | Percent Identity: | 24.000 |

alignment_block:
US-09-303-518D-569 x AAB88540 ..

Align seg 1/1 to: AAB88540 from: 1 to: 318

```
  70 CTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAA 119
     |||      |||:::::::::|||:::               |||   ::
  39 LeuLeuLeuLeuAlaPheValProPheArgLeuArgAspLysLeuThrGl 55

 120 CCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCA 169
     ::::||||||            :::      |||:::::|||::|||
  55 yLysLeuGlyIleTrpIleGlyHisLysAlaLysLysGlnArgThrArgA 72

 170 TCGTCGCCAATATGCGTCAGGCA.....GGCATGAATCCCGAC 207
           :::|||::::::   :::          :::
  72 laGlnThrAsnLeuGlnTyrCysPheProHisTrpThrGluGlnGlnArg 88

 208 CCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGA 257
     :::::::::      :::||||||   :::|||:::      :::
  89 GluGlnValIleAspLysMetPheAlaValValAlaGlnValMetPheGl 105

 258 ACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGTTCA 307
      ::::::         :::|||:::   :::   ::::::         :
 105 yIleGlyGluIleAlaIleArgSerLysLysHisLeuGlnLysArgSerG 122
```

```
    308 AAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAA 357
        ::   :::   |||   ||||||::::::||||||      :::   :::
    122 luPheIle...GlyLeuGluHisIleGluGlnAlaLysAlaGluGlyLys 137

    358 GGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACG 407
        ::::::::::::::::::||||||      :::   |||   :::|||
    138 AsnIleIleLeuMetValProHisGlyTrpAlaIleAspAlaSerGlyIl 154

    408 CTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGC 457
            :::      |||      :::|||:::|||:::||||||:::|||
    154 eIleLeuHisThrGln...GlyMetProMetThrSerMetTyrAsnProH 170

    458 CGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGC 507
          :::         :::|||   :::             |||   |||
    170 isArgAsnProLeuValAspTrpLeuTrpThrIleThrArgGlnArgPhe 186

    508 AAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAA 557
          ||||||              :::   :::|||:::|||   ::::::::
    187 GlyGlyLysMetHisAlaArgGln...AsnGlyIleLysProPheLeuSe 202

    558 AGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCT 607
        :   :::|||:::||||||           |||||||||
    202 rHisValArgLysGlyGluMetGlyTyrTyrLeuProAsp..... 215

    608 CCCCTCAAGAAGGCGGG......GAAGGCGTATGGGTGGATTTCTTCGGC 651
            ::::::   |||      ::::::|||:::|||||||||||||||
    216 .....GluAspPheGlyAlaGluGlnSerValPheValAspPhePheGly 230

    652 AAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTC...AA 698
                  |||:::            |||:::|||   :::   ||
    231 ThrTyrLysAlaThrLeuProGlyLeuAsnLysMetAlaLysLeuSerLy 247

    699 AGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAG 748
        |:::|||         :::|||         :::       |||:::
    247 sAlaValValIleProMetPheProArgTyrAsnAlaGluThrGlyLys. 263

    749 GTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAA 798
          :::::::::::::|||:::|||::::::   :::|||:::   |||
    264 ..TyrGluMetGluIleHisProAlaMet...AsnLeuSerAspAspPro 278

    799 GCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCG 848
          ::::::|||   ::::::|||::::::   |||   ::::::
    279 GluGlnSerAlaArgAlaMetAsnGluGluIleGluSerPheValThrPr 295

    849 TTTTCCGACGCAGTATCTGTTTATG 873
            |||   ||||||::::::::::
    295 oAlaProGluGlnTyrValTrpIle 303

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60654

seq_documentation_block:
ID   AAB60654 standard; Protein; 318 AA.
XX
AC   AAB60654;
XX
DT   04-MAY-2001  (first entry)
XX
DE   Haemophilus influenzae MsbB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; MsbB.
XX
OS   Haemophilus influenzae.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91453.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 98-99; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Haemophilus
CC   influenzae MsbB protein.
XX
SQ   Sequence   318 AA;


alignment_scores:
            Quality:  178.50            Length:    275
              Ratio:  1.137               Gaps:      9
 Percent Similarity:  57.091   Percent Identity: 24.000

alignment_block:
 US-09-303-518D-569 x AAB60654   ..

 Align seg 1/1  to: AAB60654  from: 1  to: 318

     70 CTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAA 119
        |||      |||::::::::::|||:::                |||   ::
     39 LeuLeuLeuLeuAlaPheValProPheArgLeuArgAspLysLeuThrGl 55

    120 CCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCA 169
        :::|||||||           :::      |||:::::::|||:::|||
     55 yLysLeuGlyIleTrpIleGlyHisLysAlaLysLysGlnArgThrArgA 72

    170 TCGTCGCCAATATGCGTCAGGCA.....GGCATGAATCCCGAC 207
            :::|||::::::   :::                 :::
     72 laGlnThrAsnLeuGlnTyrCysPheProHisTrpThrGluGlnGlnArg 88

    208 CCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGA 257
          :::::::::      :::||||||   :::|||:::      :::
     89 GluGlnValIleAspLysMetPheAlaValValAlaGlnValMetPheGl 105
```

```
 258 ACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGTTCA 307
     ::::::         :::|||:::   :::   ::::::         :
 105 yIleGlyGluIleAlaIleArgSerLysLysHisLeuGlnLysArgSerG 122

 308 AAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAA 357
     ::   :::   |||   ||||||::::::||||||      :::   :::
 122 luPheIle...GlyLeuGluHisIleGluGlnAlaLysAlaGluGlyLys 137

 358 GGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACG 407
     ::::::::::::::::::||||||      :::   |||   :::|||
 138 AsnIleIleLeuMetValProHisGlyTrpAlaIleAspAlaSerGlyIl 154

 408 CTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGC 457
         :::      |||      :::|||:::|||:::||||||:::|||
 154 eIleLeuHisThrGln...GlyMetProMetThrSerMetTyrAsnProH 170

 458 CGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGC 507
       :::         :::|||   :::             |||   ||| ,
 170 isArgAsnProLeuValAspTrpLeuTrpThrIleThrArgGlnArgPhe 186

 508 AAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAA 557
       ||||||            :::   :::|||:::|||   ::::::::
 187 GlyGlyLysMetHisAlaArgGln...AsnGlyIleLysProPheLeuSe 202

 558 AGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCT 607
     :   :::|||:::||||||            ||||||||
 202 rHisValArgLysGlyGluMetGlyTyrTyrLeuProAsp......... 215

 608 CCCCTCAAGAAGGCGGG......GAAGGCGTATGGGTGGATTTCTTCGGC 651
         ::::::   |||      ::::::|||:::||||||||||||||||
 216 .....GluAspPheGlyAlaGluGlnSerValPheValAspPhePheGly 230

 652 AAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTC...AA 698
                |||:::           |||:::|||   :::   ||
 231 ThrTyrLysAlaThrLeuProGlyLeuAsnLysMetAlaLysLeuSerLy 247

 699 AGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAG 748
     |:::|||         :::|||         :::         |||:::
 247 sAlaValValIleProMetPheProArgTyrAsnAlaGluThrGlyLys. 263

 749 GTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAA 798
       ::::::::::::|||:::|||::::::   :::|||:::   |||
 264 ..TyrGluMetGluIleHisProAlaMet...AsnLeuSerAspAspPro 278

 799 GCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCG 848
       ::::::|||   ::::::|||::::::   |||   ::::::
 279 GluGlnSerAlaArgAlaMetAsnGluGluIleGluSerPheValThrPr 295

 849 TTTTCCGACGCAGTATCTGTTTATG 873
       |||   ||||||:::::::::
 295 oAlaProGluGlnTyrValTrpIle 303
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60656

seq_documentation_block:

```
ID   AAB60656 standard; Protein; 291 AA.
XX
AC   AAB60656;
XX
DT   04-MAY-2001  (first entry)
XX
DE   N. meningitidis (serogroup B) MsbB protein.
XX
KW   Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW   genetically modified; protective antigen expression; LPS detoxification;
KW   LPS; lipid A; homologous recombination vector; immunisation;
KW   immunoprotective; non-toxic; paediatric; MsbB.
XX
OS   Neisseria meningitidis.
XX
PN   WO200109350-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07424.
XX
PR   03-AUG-1999;   99GB-0018319.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Berthet FJ,  Dalemans WLJ,  Denoel P,  Dequesne G,  Feron C,  Lobet Y;
PI   Poolman J,  Thiry G,  Thonnard J,  Voet P;
XX
DR   WPI; 2001-138654/14.
DR   N-PSDB; AAF91455.
XX
PT   New isolated polynucleotide useful for outer membrane vesicle
PT   preparation from Gram-negative bacterial strain for vaccination of
PT   microbial infections -
XX
PS   Disclosure; Page 99; 128pp; English.
XX
CC   The invention relates to a genetically-engineered outer membrane vesicle
CC   (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC   The blebs of the invention are improved with respect to their
CC   immunogenicity and toxicity by the introduction of one or more genetic
CC   changes to the chromosome of the bacterium  from which the blebs are
CC   derived. The changes made include the upregulation of protective antigen
CC   expression, the downregulation of immunodominant non-protective antigen
CC   expression, and genetic changes which result in detoxification of the
CC   Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC   encompasses modified Gram-negative bacterial strains from which the bleb
CC   preparations are made, a vector suitable for performing recombination
CC   events (for the generation of the modified bacterial strains),
CC   bacterially-derived nucleic acid sequences used in such a vector, and an
CC   immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC   cell vaccine suitable for paediatric use. The bleb preparation is useful
CC   in the manufacture of a medicament for immunising a human host against a
CC   disease caused by infection of one or more of the following:  Neisseria
CC   meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC   catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC   pneumonia. The invention may also be used to provide immunisation against
CC   the influenza virus. Bacterially derived nucleotide sequences of the
CC   invention are used in the performance of homologous recombination events
CC   up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC   increase or decrease expression of that gene. Immunoprotective and
CC   non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC   are more immunogenic, less toxic and safer, and are particularly useful
CC   for paediatric use. The present sequence represents Neisseria
CC   meningitidis MsbB protein.
XX
SQ   Sequence   291 AA;
```

alignment_scores:

```
           Quality:  174.50              Length:    286
             Ratio:  1.064                 Gaps:     10
 Percent Similarity:  57.343   Percent Identity: 25.874
```

alignment_block:

US-09-303-518D-569 x AAB60656 ..

Align seg 1/1 to: AAB60656 from: 1 to: 291

```
  67 CTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGG 116
     :::|||      |||::::::||||||::::::   ||||||   |||::
   7 ValLeuTyrValLeuGlnPheLeuProPheAlaLeuLeuHisLysLeuAl 23

 117 AAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGC 166
     ::::      |||   |||||||:::   |||:::|||      |||
  23 aAspLeuThrGlyLeuLeuAlaTyrLeuLeuValLysProArgArgArgI 40

 167 GCATCGTCGCCAATATGCGTCAGGCAGGCATGAATCCCGACCCCAAA... 213
               |||:::   ::::::      :::   |||   |||
  40 leGlyGluIleAsnLeuAlaLysCysPheProGluTrpAspGlyLysLys 56
```

```
214 .........ACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTTT 254
              :::::::|||      |||   :::   ||||||      ||
 57 ArgGluThrValLeuLysGlnHisPheLysHisMetAlaLysLeuMetLe 73

255 GGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGT 304
    ||||   :::      ::::::                ::::::::::::
 73 uGluTyrGlyLeuTyrTrpTyrAlaProAlaGlyArgLeuLysSerLeuV 90

305 TCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACAC 354
      :::      ::::::   ::::::::::::::::||||||
 90 alArg...TyrArgAsnLysHisTyrLeuAspAspAlaLeuAlaAlaGly 105

355 GAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGG 404
    |||   ::::::::::::   ||||||:::   ::::::::::::::
106 GluLysValIleIleLeuTyrProHisPheThrAlaPheGluMetAlaVa 122

405 ACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAAC 454
          :::::::|||::::::      ||||||   :::|||||||:::
122 lTyrAlaLeuAsnGlnAspVal......ProLeuIleSerMetTyrSerH 137

455 CGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGC 504
        |||   |||   :::|||      :::      ||||||   |||
137 isGlnLysAsnLysIleLeuAspAlaGlnIleLeuLysGlyArgAsnArg 153

505 GGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCAT 554
          :::                  :::||||||:::   :::::
154 TyrAspAsnValPheLeuIleGlyArgThrGluGlyValArgAlaLeuVa 170

555 CAAAGCCCTGCGTTCGGGCGAAGCAACC...ATCGTCCTGCCCGACCACG 601
    :|||   :::|||:::::::::|||      :::   |||||||||
170 lLysGlnPheArgLysSerSerAlaProPheLeuTyrLeuProAsp.... 185

602 TCCCCTCCCCTCAAGAAGGCGGG......GAAGGCGTATGGGTGGATTTC 645
              |||:::   |||      ::::::|||:::||||||||
186 ...........GlnAspPheGlyArgAsnAspSerValPheValAspPhe 198

646 TTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGT 695
    ||||||      :::   |||:::|||      ::::::::::|||   ::
199 PheGlyIleGlnThrAlaThrIleThrGlyLeuSerArgIleAlaAlaLe 215

696 CAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCT.......738
    :   ::::::|||::::::                  :::|||
215 uAlaAsnAlaLysValIle............ProAlaIleProValArgG 228

739 ..GGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTG 786
      :::   :::        ||||||:::   |||:::      ::::::
228 luAlaAspAsnThrValThrLeuHisPheTyrProAlaTrpGluSerPhe 244

787 AACGGCGACAAAGCCCAT...GATGCCGCCGTGTTCAACCGCAATGCCGA 833
      ::::::   |||:::   ||||||      :::||||||      ||
245 ProSerGluAspAlaGlnAlaAspAlaGlnArgMetAsnArgPheIleGl 261

834 ATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCT 883
    |           ::::::|||::: ||||||      ||||||::::::
261 uGluProCysAlaAsnIleProSer.SerIlePheGlyCysThrSerVal 277

884 ACAAAA 889
    :::|||
278 SerLys 279
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37217

seq_documentation_block:

ID AAY37217 standard; Protein; 455 AA.
XX
AC AAY37217;
XX
DT 07-OCT-1999 (first entry)
XX
DE Amino acid sequence of a Chlamydia trachomatis protein.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST ) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Disclosure; Page 981-982; 1755pp; English.
XX
CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
SQ Sequence 455 AA;

alignment_scores:

```
            Quality:  172.50              Length:    342
              Ratio:  1.033                 Gaps:     12
 Percent Similarity:  48.830   Percent Identity: 23.977
```

alignment_block:
US-09-303-518D-569 x AAY37217 ..

Align seg 1/1 to: AAY37217 from: 1 to: 455

```
 22 CTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTG.. 69
    ||||||      ||||||||||:::      ||||||:::      |||
  2 LeuPheLysArgLeuArgThrGlyGlyLysIleLeuValAspHisLeuVa 18

 70 .....................CTCAAATGCCTCTCCCTGCTGCCGCTTT 97
                           |||      |||   |||||||||   |
 18 lTyrGlyLeuGlyLeuGlyValLeuThrIleLeuArgLeuLeuProArgS 35

 98 CCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTT 147
    ||   |||:::   :::::::::   ||||||          ||||||:::
 35 erSerLeuArgLeuPheSerLysGlyLeuGlyThrAlaLeuPheTyrPhe 51

148 TTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATG..............  183
    :::::::::   |||         :::::::|||:::
 52 IleSerAspPheArgLysThrAlaLeuThrAsnLeuAlaLeuAlaPhePr 68

184 ................................CGTCAGGCA.......  192
                                    |||||||:::
 68 oGluLysSerPheAlaGluArgTyrGlnIleAlaArgGlnSerValGlnG 85

193 ...................GGCATGAATCCCGACCCCAAAACGGTCAAA 222
                       :::            |||   :::|||
 85 lnMetIleIleThrPheValGluLeuAlaThrValAspLysPheAlaLys 101
```

```
 223 GCCGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCC...GC 269
     :::            |||   :::      |||   ||||||   ::
 102 HisIleAspGluMetIleAlaIleAlaThrSerGluAspAlaProGluGl 118

 270 GTTTTTCAGAAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTACACG 319
     :||||||      ||||||:::::::::::::::   :::
 118 yPhePhe......ProGluGluValSerSerGlnGlnGluLeu..... 130

 320 GCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTATTC 369
         :::|||      :::   ||||||::::::|||||| . ::::::
 131 .....AspHisPhePheSerArgLeuAspArgGlnGluGlyAlaIleLeu 145

 370 ATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCA 419
     :::    |||   :::::::::::::|||         ||||||:::::
 146 PheCysGlyHisGlnAlaAsnTrpGluLeuProPheLeuTyrIleThrLy 162

 420 GCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAG 469
     ::::      :::|||      |||:::   ||||||   |||   :::
 162 sArg......TyrProGlyLeuAlaPheAlaLysProValLysAsnArgA 177

 470 CGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACC 519
       :::::::::   :::   :::   |||          :::||||||
 177 rgLeuAsnGlnLysIleIleSerLeuArgGluSerPheGlnGlyLysIle 193

 520 GCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTC 569
     :::|||   :::   :::::::::::|||   ::::::||||||:::
 194 ValProProGln...AsnAlaIleAsnGlnAlaLeuArgAlaLeuHisAr 209

 570 GGGCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCTCCCCTCAAGAAG 619
      ||||||::::::   ::::::   |||:::|||         :::|||
 209 gGlyGluValValGlyIleValGlyAspGlnValLeuLeuSerSerGluT 226

 620 GCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATACCATG 669
       :::                 :::||||||:::   |||:::|||
 226 yrSer............TyrProLeuPheGlySerGlnAlaPheThrThr 238

 670 ACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTT 719
     |||      |||   ||||||:::                 ::::::
 239 ThrSerProAlaLeuLeuAlaTyrLysThrLysLysThrValIleAlaVa 255

 720 CTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCC 769
      :::      |||   |||:::|||:::   :::   :::
 255 lAlaIleTyrArgLysProAsnGlyAsnTyrLeuValVal..... 268

 770 CCGTCCAAGGGGAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTC 819
                                          :::   :::|||
 269 ..................................ProSerLysAlaPhe 273

 820 AACCGCAATGCCGAATATTGGATACGC..................... 846
     :::   |||:::|||      ||||||
 274 HisAlaAsnThrGluLeuSerIleArgGluSerThrGluGlnLeuMetAs 290

 847 .........CGTTTT......................CCGACGCAGT 862
             ||||||                      |||   |||:
 290 pArgLeuMetArgPheLeuGluLysGlyIleThrCysLysProGluGlnT 307

 863 ATCTGTTTATGTACAACCGCTACAAA 888
     ::|||:::::::::::::|||:::|||
 307 rpLeuTrpLeuHisLysArgTrpLys 315
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60655

seq_documentation_block:

ID AAB60655 standard; Protein; 347 AA.
XX
AC AAB60655;
XX
DT 04-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis MsbB protein.
XX
KW Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW genetically modified; protective antigen expression; LPS detoxification;
KW LPS; lipid A; homologous recombination vector; immunisation;
KW immunoprotective; non-toxic; paediatric; MsbB.
XX
OS Moraxella catarrhalis.
XX
PN WO200109350-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;
PI Poolman J, Thiry G, Thonnard J, Voet P;
XX
DR WPI; 2001-138654/14.
DR N-PSDB; AAF91454.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
XX
PS Disclosure; Page 99; 128pp; English.
XX
CC The invention relates to a genetically-engineered outer membrane vesicle
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified Gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events (for the generation of the modified bacterial strains),
CC bacterially-derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: Neisseria
CC meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC pneumonia. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events
CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC increase or decrease expression of that gene. Immunoprotective and
CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC are more immunogenic, less toxic and safer, and are particularly useful
CC for paediatric use. The present sequence represents Moraxella catarrhalis
CC MsbB protein.
XX
SQ Sequence 347 AA;

alignment_scores:

```
          Quality:  170.50              Length:    297
            Ratio:  1.072                 Gaps:     11
Percent Similarity:  53.535   Percent Identity: 26.263
```

alignment_block:

US-09-303-518D-569 x AAB60655 ..

Align seg 1/1 to: AAB60655 from: 1 to: 347

```
  55 CTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCT 104
     |||      ||||||:::      |||   :::||||||||||
  52 LeuAlaPheAlaLeuIleLeuProLeuIlePheLeuProLeuArgTrpGl 68
```

```
  105 GCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGG 154
          :::|||:::|||||||||   |||:::   ||||||   |||:
   68 nPheTrpIleGlyLysArgLeuGlyIleLeuValHisTyrLeuAlaLysS 85

  155 AAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAATCCC 204
      ::                :::::::|||::::::   :::      |||
   85 erArgValGlnAspThrLeuThrAsnLeuGlnLeuThrPheProAsn... 100

  205 GACCCCAAA..............ACGGTCAAAGCCGTTTTTGCGGAAAC 239
      :::||||||              |||::::::   ||||||   :::
  101 GlnProLysSerLysHisLysAlaThrAlaArgGlnValPheIleAsnGl 117

  240 GGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAG 289
      :::   |||   :::|||         |||:::||||||   |||:::
  117 nGlyIleGlyIlePheGluSerLeuCysAlaTrpPheArg...ProAsnV 133

  290 ACATAGAAACAATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAG 339
        ::::::      |||   ::::::   |||   :::|||:::   :::
  133 alPheLysArgThrPhe...SerIleSerGlyLeuGlnHisLeuIleAsp 148

  340 GCTTTGGACAAACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAG 389
      |||      :::::::::::::::::::::::      |||       ::
  149 AlaGlnLysGlnAsnLysAlaValIleLeuLeuGlyGlyHisArgThrTh 165

  390 CTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGA 439
      :   ||||||||||||||      :::|||         |||
  165 rLeuAspLeuGlyGlyArgLeuCysThrGlnPhe......PheAlaAlaA 180

  440 CCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAG 489
        ::::::|||:::|||   :::         ::::::   ::::::
  180 spCysValTyrArgProGlnAsnAsnProLeuLeuGluTrpPheIleTyr 196

  490 GCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGG 539
            :::::::|||                              |||   :::
  197 ......AsnAlaArgArgCysIlePheAspGluGlnIleSerAsnArgAs 211

  540 GGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCC 589
       :::|||:::::::|||      |||:::::::|||:::
  211 pMetLysLysLeuIleThrArgLeuLysGlnGlyArgIleIleTrpTyrS 228

  590 TGCCCGACCACGTCCCCTCCCCTCAAGAAGGCGGG......GAAGGCGTA 633
        ||||||               |||:::   |||      :::||||||
  228 erProAsp..............GlnAspPheGlyLeuGluHisGlyVal 239

  634 TGGGTGGATTTCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAA 683
         :::   |||||||||   ||||||   |||:::|||          ::
  240 MetAlaThrPhePheGlyValProAlaAlaThrIleThrAlaGlnArgAr 256

  684 ATTGGCACACGTC......AAAGGCGTGAAAACCCTGTTTTTCTGCTGCG 727
      :|||     :::      |||:::         :::::::::      :
  256 gLeuIleLysLeuGlyAspLysAlaAsnProProValIleIleMetMetA 273

  728 AACGCCTG.....................CCTGGCGGACAAGGTTTCGAT 756
      ::   |||                        |||:::
  273 spMetLeuArgGlnThrProAspTyrIleAlaLysGlyHisArgProHis 289

  757 TTGCACATC.........CGCCCCGTCCAAGGGGAATTGAACGGCGACAA 797
         ||||||               :::::::::      :::   ||||::
  290 TyrHisIleSerLeuSerAlaValLeuLysAsnTyrProSerAspAspGl 306

  798 AGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCC 847
      ::::   ||||||      :::||||||      |||      |||::::
  306 uThrAlaAspAlaGluArgIleAsnArgLeuIleGluGlnAsnIleGlnL 323

  848 GTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAA 888
      ::      ||||||:::::::::::::::::::|||:::|||
  323 ysAspLeuThrGlnTrpMetTrpPheHisArgArgPheLys 336
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB60653

seq_documentation_block:
ID AAB60653 standard; Protein; 311 AA.

XX
AC AAB60653;
XX
DT 04-MAY-2001 (first entry)
XX
DE Haemophilus influenzae HtrB protein.
XX
KW Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW genetically modified; protective antigen expression; LPS detoxification;
KW LPS; lipid A; homologous recombination vector; immunisation;
KW immunoprotective; non-toxic; paediatric; HtrB.
XX
OS Haemophilus influenzae.
XX
PN WO200109350-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;
PI Poolman J, Thiry G, Thonnard J, Voet P;
XX
DR WPI; 2001-138654/14.
DR N-PSDB; AAF91452.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
XX
PS Disclosure; Page 98; 128pp; English.
XX
CC The invention relates to a genetically-engineered outer membrane vesicle
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified Gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events (for the generation of the modified bacterial strains),
CC bacterially-derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: Neisseria
CC meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC pneumonia. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events
CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC increase or decrease expression of that gene. Immunoprotective and
CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC are more immunogenic, less toxic and safer, and are particularly useful
CC for paediatric use. The present sequence represents Haemophilus
CC influenzae HtrB protein.
XX
SQ Sequence 311 AA;

alignment_scores:
Quality: 148.50 Length: 294
Ratio: 0.958 Gaps: 11
Percent Similarity: 52.721 Percent Identity: 24.490

```
alignment_block:
 US-09-303-518D-569 x AAB60653   ..

 Align seg 1/1  to: AAB60653  from: 1  to: 311

      64 GCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCT 113
         |||:::   :::   :::      ||||||         |||:::   ::
      26 AlaIleTrpArgSerIleLeuCysLeuProTyrProIleLeuArgHisIl 42

     114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
         :|||:::   :::|||   |||      :::|||             |||
      42 eGlyHisGlyPheGlyTrpLeuPheSerHisLeuLysValGlyLysArgA 59

     164 CGCGCATCGTCGCC................................... 177
                 :::|||
      59 rgAlaAlaIleAlaArgArgAsnLeuGluLeuCysPheProAspMetPro 75

     178 ...AATATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGC 224
            |||   |||::::::   ::::::   :::   ::::::|||
      76 GluAsnGluArgGluThrIleLeuGlnGluAsnLeuArgSerValGlyMe 92

     225 CGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTT 274
          ::::::   ||||||:::                          |||:::|
      92 tAlaIleIleGluThrGlyMet.....................AlaTrpP 102

     275 TCAGAAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTACACGGCTGG 324
         ||   :::   :::   |||:::          |||   ||||:::|||
     102 heTrpSerAspSerArgIleLysLysTrpSerLys...ValGluGlyLeu 117

     325 GAACATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTATTCATCAC 374
         ::::::::::::::      ::::::   :::|||:::::::::::::
     118 HisTyrLeuLysGlu......AsnGlnLysAspGlyIleValLeuValGl 132

     375 GCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGC 424
           |||:::   :::   :::|||||||:::|||   |||:::      |
     132 yValHisPheLeuThrLeuGluLeuGlyAlaArgIleIleGly......L 147

     425 TTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATA 474
         ||      |||      ::::::|||:::|||                :::
     147 euHisHisProGlyIleGlyValTyrArgProAsnAspAsnProLeuLeu 163

     475 GACAAAATCATGCAGGCGGGCAGGGTTCGC...GGCAAAGGAAAAACCGC 521
         |||   ::::::      ||||||:::|||   :::|||
     164 AspTrpLeuGlnThrGlnGlyArgLeuArgSerAsnLysAspMetLeuAs 180

     522 GCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGG 571
               :::::::::|||          :::|||||||||||||||
     180 pArgLysAspLeuArgGly.........MetIleLysAlaLeuArgHisG 194

     572 GCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCTCCCCTCAAGAAGGC 621
         |||:::               |||||||||              :::
     194 luGluThrIleTrpTyrAlaProAspHis..............AspTyr 205

     622 GGG......GAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATAC 665
         |||      ::::::|||:::|||   |||||||:::   |||
     206 GlyArgLysAsnAlaValPheValProPhePheAlaValProAspThrCy 222

     666 CATGACGCTGGCGGCA......AAATTGGCACACGTCAAAGGCGTGAAAA 709
            |||   ::::::          |||          ::::::  |||:
     222 sThrThrThrGlySerTyrTyrLeuLeuLysSerSerGlnAsnSerLysV 239

     710 CCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTG 759
         :::::   |||:::      |||        |||:::|||:::   :::
     239 alIleProPheAlaProLeuArgAsnLysAspGlySerGlyTyrThrVal 255

     760 CACATC...CGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGCCCATGA 806
            |||      ||||||:::                 |||::::::
     256 SerIleSerAlaProValAspPheThrAspLeuGlnAspGluThrAlaIl 272

     807 TGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTTTTCCGA 856
         ||||||   :::|||:::   :::|||      |||   :::      :
     272 eAlaAlaArgMetAsnGlnIleValGluLysGluIleMetLysGlyIleS 289

      857 CGCAGTATCTGTTTATGTACAACCGCTACAAA 888
          ::|||||||:::::::::::::::::|||:::|||
      289 erGlnTyrMetTrpLeuHisArgArgPheLys 299

 seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34697

 seq_documentation_block:
 ID   AAY34697 standard; Protein; 463 AA.
 XX
 AC   AAY34697;
 XX
 DT   13-SEP-1999  (first entry)
 XX
 DE   Chlamydia pneumoniae lipoprotein sequence.
 XX
 KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW   vaccine; neutralising epitope.
 XX
 OS   Chlamydia pneumoniae.
 XX
 PN   WO9927105-A2.
 XX
 PD   03-JUN-1999.
 XX
 PF   20-NOV-1998;   98WO-IB01890.
 XX
 PR   04-NOV-1998;   98US-0107078.
 PR   21-NOV-1997;   97FR-0014673.
 XX
 PA   (GEST ) GENSET.
 XX
 PI   Griffais R;
 XX
 DR   WPI; 1999-357842/30.
 XX
 PT   Genome sequence of Chlamydia pneumoniae
 XX
 PS   Page 699-700; Disclosure; 1912pp; English.
 XX
 CC   AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC   frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC   C. pneumoniae causes respiratory disease such as pneumonia and
 CC   bronchitis and is thought to be a contributing factor in heart
 CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC   frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC   nucleotides sequences can also be used as immunogenic compositions,
 CC   especially where the vector directs the expression of a neutralising
 CC   epitope of C. pneumoniae.
 XX
 SQ   Sequence   463 AA;


alignment_scores:
            Quality:  146.00              Length:    331
              Ratio:  0.869                Gaps:     13
 Percent Similarity:  50.755   Percent Identity: 22.356

alignment_block:
 US-09-303-518D-569 x AAY34697   ..

 Align seg 1/1  to: AAY34697  from: 1  to: 463

      34 TTGCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTC 83
         |||:::::   ::::::   |||:::::::::::::::
      14 LeuGluAlaProLeuTyrTyrLeuValSerGlyIleIleAlaLeuCysAr 30

      84 CCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATC 133
                |||   |||   |||      |||||||:::   :::|||   |
      30 gHisThrProArgSerPheLeuThrGlyLeuGlyLysGlyPheGlyPheL 47
```

```
 134 TGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATG 183
     ||||||||||||:::::::::::::   |||         :::::::|||:::
  47 euAlaPheTyrIleIleSerAspTyrArgLysThrAlaLeuThrAsnLeu 63

 184 ..................................................CG 185
                                                        ||
  64 AlaLeuAlaPheProGluLysThrPheAspGluArgTyrLysIleAlaAr 80

 186 TCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGG 235
     ||||:::   ::::::                 ::::::   :::       |
  80 gGlnSerLeuGlnHisLeuIleIleThrLeuLeuGluLeuAlaIleG 97

 236 AAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTC..... 276
     ||         |||:::::::::         ::::::
  97 luGlnLeuValGlyAsnIleAspLysLeuIleThrIleValThrSerSer 113

 277 AGAAAACCG..............................GAAGACAT 293
     |||:::|||                              |||||||::
 114 ArgAsnProLysGlyPheSerSerGluGluValIleSerAsnGluAspLe 130

 294 AGAAACAATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTT 343
     :|||      ||||||                                 |
 130 uGluGluThrPheLys.............................AsnL 137

 344 TGGACAAACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTAC 393
     ||::::::   :::|||||||:::::::::      |||   :::::::::
 137 euGlnGluLysGlnGlyLeuIleLeuPheCysGlyHisGlnAlaAsnTrp 153

 394 GATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGC 443
     :::|||         ||||||::::::::      :::|||      ||
 154 GluLeuProPheLeuTyrIleThrLysAsn......TyrProGlyIleAl 168

 444 CATGTACAAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGG 493
     |:::   |||      |||   :::   ::::::|||   :::   |||
 168 aPheAlaLysAlaIleLysAsnGlnArgLeuSerLysLysIlePheAlaL 185

 494 GCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTC 543
       |||          |||||||||   :::|||   :::   :::|||:::
 185 euArgGluValPheLysGlyLysIleValProProLys...AsnGlyIle 200

 544 AAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCC 593
     :::|||   |||:::|||||||:::::|||:::   :::   ::::::
 201 GlnGlnGlyIleGluAlaLeuAsnGlnGlyLysLeuValGlyIleValGl 217

 594 CGACCACGTC.....................CCCTCCCCTCAAGAAG 619
      |||::::::                     |||
 217 yAspGlnAlaLeuLeuMetSerSerTyrThrTyrPro............. 229

 620 GCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATACCATG 669
                        :::|||||||:::|||||||:::|||
 230 .....................LeuPheGlySerProAlaPheThrThr 238

 670 ACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTT 719
     |||      |||   ||||||:::      |||      ::::::
 239 ThrSerProAlaLeuLeuAlaTyrLysThrGlyPheProValIleAlaVa 255

 720 CTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCC 769
           :::|||         ::::::|||||||:::      :::   |
 255 lAsnValSerArg......GlnAlaLysGlyPheGlu......ValIleP 268

 770 CCGTCCAAGGGGAATTGAACGGCGACAAAGCC............CATGAT 807
     ||   :::::::::|||   :::::::|||:::           ::::::
 268 ro...SerAlaLysLeuTyrAlaAsnLysSerLeuProMetLysGluSer 283

 808 GCCGCCGTGTTCAACCGCAAT............GCCGAATATTGGATACG 845
     :::|||::::::      :::                |||      |||
 284 ValAlaIleLeuMetAspGlnMetMetGlyPheLeuGluLysGlyIleAl 300

 846 CCGTTTTCCGACGCAGTATCTGTTTATGTACAACCGCTACAAA 888
           |||   |||:::::::::::::::::::|||:::|||
 300 aSerGlnProGluGlnTrpMetTrpIleHisLysArgTrpLys 314
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW25084

```
seq_documentation_block:
ID   AAW25084 standard; Protein; 311 AA.
XX
AC   AAW25084;
XX
DT   30-DEC-1997  (first entry)
XX
DE   Haemophilus influenzae htrB polypeptide.
XX
KW   Vaccine; htrB gene; Gram-negative bacterium; non-toxic mutant;
KW   pathogen; endotoxin; diagnosis; passive immunisation.
XX
OS   Haemophilus influenzae strain 2019.
XX
PN   WO9719688-A1.
XX
PD   05-JUN-1997.
XX
PF   27-NOV-1996;   96WO-US18984.
XX
PR   01-DEC-1995;   95US-0565943.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (REGC ) UNIV CALIFORNIA.
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Apicella MA,  Arumugham R,  Gibson BW,  Lee N,  Sunshine MG;
XX
DR   WPI; 1997-310355/28.
DR   N-PSDB; AAT79708.
XX
PT   New Gram-negative bacterial pathogen vaccines - comprising a htrB
PT   mutant or an endotoxin isolated from an htrB mutant optionally
PT   conjugated to a carrier protein.
XX
PS   Example 1; Page 61-62; 79pp; English.
XX
CC   This polypeptide comprises the htrB gene product (see also AAT79708)
CC   of Haemophilus influenzae strain 2019.  A claimed vaccine
CC   formulation contains as an active ingredient an htrB mutant of a
CC   Gram-negative bacterial pathogen (GNBP), endotoxin isolated from an
CC   htrB mutant (A) of a GNBP, endotoxin isolated from (A) conjugated
CC   to a carrier protein, or (A) which has been genetically engineered
CC   to express at least one heterologous vaccine antigen, where (A)
CC   lacks one or more secondary acyl chains of lipid A contained in the
CC   GNBP resulting in reduced toxicity when compared to lipid A of the
CC   GNBP.  Also claimed is a method for producing endotoxin-specific
CC   antisera for diagnostic assays, or for passive immunisation,
CC   comprising immunising an individual with a vaccine formulation
CC   comprising an active ingredient as above, and collecting antibodies
CC   produced from the immunised individual.
XX
SQ   Sequence   311 AA;


alignment_scores:
            Quality:  144.50            Length:    294
              Ratio:  0.932               Gaps:     11
 Percent Similarity:  52.721   Percent Identity: 24.150

alignment_block:
 US-09-303-518D-569 x AAW25084   ..

 Align seg 1/1  to: AAW25084  from: 1  to: 311

      64 GCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCT 113
         |||:::   :::   :::      ||||||         |||:::   ::
      26 AlaIleTrpArgSerIleLeuCysLeuProTyrProIleLeuArgHisIl 42

     114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
         :|||:::   :::|||   |||      :::|||            |||
      42 eGlyHisGlyPheGlyTrpLeuPheSerHisLeuLysValGlyLysArgA 59
```

```
 164 CGCGCATCGTCGCC..................................... 177
                ::::|||
  59 rgAlaAlaIleAlaArgArgAsnLeuGluLeuCysPheProAspMetPro 75

 178 ...AATATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGC 224
        |||   |||::::::   ::::::   :::   :::::::|||
  76 GluAsnGluArgGluThrIleLeuGlnGluAsnLeuArgSerValGlyMe 92

 225 CGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTT 274
      ::::::   ||||||:::                          |||:::|
  92 tAlaIleIleGluThrGlyMet.....................AlaTrpP 102

 275 TCAGAAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTACACGGCTGG 324
     ||   :::   :::   |||:::          |||   |||:::|||
 102 heTrpSerAspSerArgIleLysLysTrpSerLys...ValGluGlyLeu 117

 325 GAACATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTATTCATCAC 374
     ::::::::::::::      ::::::   :::|||::::::::::::
 118 HisTyrLeuLysGlu......AsnGlnLysAspGlyIleValLeuValGl 132

 375 GCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGC 424
       |||:::   :::   :::||||||:::|||   |||:::      |
 132 yValHisPheLeuThrLeuGluLeuGlyAlaArgIleIleGly......L 147

 425 TTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATA 474
     ||      |||      ::::::|||:::|||                :::
 147 euHisHisProGlyIleGlyValTyrArgProAsnAspAsnProLeuLeu 163

 475 GACAAAATCATGCAGGCGGGCAGGGTTCGC...GGCAAAGGAAAAACCGC 521
     |||   ::::::      |||||||:::|||   :::|||
 164 AspTrpLeuGlnThrGlnGlyArgLeuArgSerAsnLysAspMetLeuAs 180

 522 GCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGG 571
           :::::::::|||          :::|||||||||||||||
 180 pArgLysAspLeuArgGly.........MetIleLysAlaLeuArgHisG 194

 572 GCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCTCCCCTCAAGAAGGC 621
       |||:::              |||||||||              :::
 194 luGluThrIleTrpTyrAlaProAspHis..............AspTyr 205

 622 GGG......GAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATAC 665
     |||      ::::::|||:::|||   ||||||:::   |||
 206 GlyArgLysAsnAlaValPheValProPhePheAlaValProAspThrCy 222

 666 CATGACGCTGGCGGCA......AAATTGGCACACGTCAAAGGCGTGAAAA 709
         |||   ::::::         |||         ::::::   |||:
 222 sThrThrThrGlySerTyrTyrLeuLeuLysSerSerGlnAsnSerLysV 239

 710 CCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTG 759
     :::::   |||:::      |||         |||:::|||:::   :::
 239 alIleProPheAlaProLeuArgAsnLysAspGlySerGlyTyrThrVal 255

 760 CACATC...CGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGCCCATGA 806
        |||      ||||||:::                  |||::::::
 256 SerIleSerAlaProValAspPheThrAspLeuGlnAspGluValAlaIl 272

 807 TGCCGCCGTGTTCAACCGCAATGCCGAATATTGGATACGCCGTTTTCCGA 856
      |||:::   :::|||:::   :::|||      |||   :::      :
 272 eAlaValArgMetAsnGlnIleValGluLysGluIleMetLysGlyIleS 289

 857 CGCAGTATCTGTTTATGTACAACCGCTACAAA 888
     ::||||||:::::::::::::::|||:::|||
 289 erGlnTyrMetTrpLeuHisArgArgPheLys 299
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59824

seq_documentation_block:

ID AAB59824 standard; Protein; 1605 AA.
XX
AC AAB59824;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #1 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1605 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 131.00 | Length: | 327 |
| Ratio: | 0.929 | Gaps: | 15 |
| Percent Similarity: | 43.119 | Percent Identity: | 25.994 |

alignment_block:
US-09-303-518D-569 x AAB59824 ..

Align seg 1/1 to: AAB59824 from: 1 to: 1605

```
  48 GCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTT 97
     |||   |||   :::||||||   |||:::      |||:::|||
 568 AlaGlyProGlyGluArgProGluGlnValAsp...ProGlyAlaProGl 583

  98 CCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTT 147
               |||         |||   |||   |||      :::|||
 583 yGlyArgIleHisArgArgLeuProArgArgArgSerGlnHisValProA 600

 148 TTAAAGGAAGACCGCGCGCGCATCGTCGCCAA..................... 179
           :::|||         |||   |||
 600 laValSerArgThrValProHisGlyArgAlaGlyLeuProAlaGluArg 616

 180 ......TATGCGTCAGGCAGGCAT.....................GAATC 202
           |||:::||||||:::                        |||::
 617 LeuLeuAlaAlaAlaGlyArgArgGlyGlyGlyAspGlnIleLeuGluAl 633

 203 CCGACCC.......................CAAAACGGTCAAAGCCGTT 228
     :   |||                      ::::::|||:::   |||
 633 aAlaProAlaGluGlnValSerAlaLeuPheArgSerGlyArgProArgP 650
```

```
 229 TTTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAG 278
650 roHisValSerGlyGlnGlnHisGlyGlyAlaValLeuArgPhe..... 664

 279 AAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTACACGGCTGGGAAC 328
665 ArgLeuGlnGlnHisArgAlaAlaLeuArgAsn.....ArgProGlyAr 679

 329 ATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTATTCATCACGCCG 378
679 gArgAlaAlaGlyAlaHisGlnAlaArgArgLysAlaTyrArgArgSerP 696

 379 CACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCC 428
696 roGlyArgHisVal.....Asp 701

 429 GTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAGACA 478
702 LeuProLeuGluArgHisGluGlySerArgGlnHisArgGlnAspArgGl 718

 479 AAATCATGCAGGCGGG.....CAGGGTTCGCGGCAA.....A 510
718 nLeuGluGlyAspGlyHisArgLeuGlnGlyGlyAspGlnLeuGlyAlaP 735

 511 GGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGC 560
735 roAlaGlyProAlaValGlnAspArgArgGlyLysLeuArgAspArgSer 751

 561 CCTGCGTTCGGGCGAAGC.....AACCATCGTCCTGCCCGACC 598
752 GluAlaProGlyArgAlaAlaArgAsnArgArgHisLeuSerAlaHisSe 768

 599 ACGTCCC..... 605
768 rArgArgAlaLeuGlnGlyProGlnGlyArgAspAlaGlyGluIleLeuT 785

 606 ..CTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAA 653
785 yrLeuProAspLeuSerArgAspArgAlaLeuArgGluArgLeuArgPro 801

 654 ACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCG 703
802 GluGlyArgHisProAlaValAlaValLeuGlnGlyLeuArgArgArgGl 818

 704 TGAAAACCCTGTTTTTCTGCTGCGAACGCCT..... 734
818 nGluIleProAla.....AspGluProHisGlyCysGlyGlyThrA 832

 735 .....GCCTGGCGGACAAGGTTTCGA 755
832 rgArgAspGlyThrPheGluAspPheArgAlaTrpArgArg.....Gln 846

 756 TTTGCACATCCGCCCCGTCCA...AGGGGAATTGAACGGCGACAAAGCCC 802
847 ValAlaArgLeuProArgAsnLeuProGlyValGluArgSerValHisPr 863

 803 ATGATGCCGCCGTGTTCAACCGCAATGCCGA 833
863 oHisArgArgArgHisGlnArgGlnGlyArg 873

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59814

seq_documentation_block:
ID   AAB59814 standard; Protein; 1022 AA.
XX
AC   AAB59814;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #5.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1022 AA;


alignment_scores:
            Quality:  130.00              Length:    321
              Ratio:  0.922                Gaps:     16
 Percent Similarity:  43.925   Percent Identity: 26.480

alignment_block:
 US-09-303-518D-569 x AAB59814   ..

 Align seg 1/1  to: AAB59814  from: 1  to: 1022

  39 AACCGCCATGCACATCCTGTTGACCGCCCT.....GCTCAAATGCC 79
  11 SerGlnHisValProAlaValSerArgThrValProHisGlyArgAlaGl 27

  80 TCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAA..... 119
  27 yLeuProAlaGluArgLeuLeuAlaAla...AlaGlyArgArgGlyGlyG 43

 120 .....CCGGCTCGGACATCTGGCGTTTTACCT 146
  43 lyAspGlnIleLeuGluAlaAlaProAlaGluGlnValSerAlaLeuPhe 59

 147 TTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCA 196
  60 ArgSerGlyArgProArgProHis.....ValSerGlyGlnGl 72

 197 TGAATCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAA 246
  72 nHis.....GlyGlyAlaValL 78

 247 GGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGA 296
  78 euArgPhe.....ArgLeuGlnGlnHisArgAlaAlaLeuArg 90

 297 AACAATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGG 346
```

```
      |||                 |||   |||      |||||||||
 91 Asn.................ArgProGlyArgArgAlaAlaGlyAlaHi 101

347 ACAAACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGAT 396
    |||:::||||||   |||   :::::::::      |||::::::
101 sGlnAlaArgArgLysAlaTyrArgArgSerProGlyArgHisVal.... 116

397 TTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCAT 446
                              :::::::|||   :::||||||
117 ............................AspLeuProLeuGluArgHis 123

447 GTACAAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGG.. 494
        :::   ::::::::::::|||||||||   ::::::   |||
124 GluGlySerArgGlnHisArgGlnAspArgGlnLeuGluGlyAspGlyHi 140

495 .......CAGGGTTCGCGGCAA.........AGGAAAAACCGCGCCTACC 528
          ||||||:::   |||              :::   |||   |
140 sArgLeuGlnGlyGlyAspGlnLeuGlyAlaProAlaGlyProAlaValG 157

529 AGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGC 578
    ||      ||||||:::   :::   :::|||   |||   |||||||:::
157 lnAspArgArgGlyLysLeuArgAspArgSerGluAlaProGlyArgAla 173

579 ............AACCATCGTCCTGCCCGACCACGTCCC............ 605
                :::|||      |||:::   |||
174 AlaArgAsnArgArgHisLeuSerAlaHisSerArgArgAlaLeuGlnGl 190

606 ................................CTCCCCTCAAGAAGGC 621
                                    ||||||:::       |
190 yProGlnGlyArgAspAlaGlyGluIleLeuTyrLeuProAspLeuSerA 207

622 GGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATACCATGAC 671
    ||   |||   :::         ||||||            :::
207 rgAspArgAlaLeuArgGluArgLeuArgProGluGlyArgHisProAla 223

672 GCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCT 721
    ::::::        |||   |||:::|||:::|||   |||:::
224 ValAlaValLeuGlnGlyLeuArgArgArgGlnGluIleProAla..... 238

722 GCTGCGAACGCCT..................................... 734
              |||
239 ....AspGluProHisGlyCysGlyGlyThrArgArgAspGlyThrPheG 254

735 ..........GCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGT 773
             |||||||||         :::   |||:::   ||||||
254 luAspPheArgAlaTrpArgArg......GlnValAlaArgLeuProArg 268

774 CCA...AGGGGAATTGAACGGCGACAAAGCCCATGATGCCGCCGTGTTCA 820
          |||:::||||||         |||      ||||||   ||
269 AsnLeuProGlyValGluArgSerValHisProHisArgArgArgHisGl 285

821 ACCGCAATGCCGA 833
    |   |||   |||
285 nArgGlnGlyArg 289
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW85665

seq_documentation_block:

ID AAW85665 standard; Protein; 323 AA.
XX
AC AAW85665;
XX
DT 19-JUL-1999 (first entry)
XX
DE Salmonella lipid A.
XX
KW Salmonella; mutant; genetic modification; msbB; lipid A; TNF;
KW tumour necrosis factor; tumour; treatment; disease; infection;
KW inhibition; melanoma; colon carcinoma; lung cancer; liver cancer;
KW kidney cancer; prostate cancer; breast cancer.
XX
OS Salmonella sp.
XX
PN WO9913053-A1.
XX
PD 18-MAR-1999.
XX
PF 09-SEP-1998; 98WO-US18701.
XX
PR 08-SEP-1998; 98US-0149832.
PR 10-SEP-1997; 97US-0926636.
XX
PA (UYYA ) UNIV YALE.
PA (VION-) VION PHARM INC.
XX
PI Bermudes D, Ittensohn M, Low KB;
XX
DR WPI; 1999-229231/19.
DR N-PSDB; AAX08503.
XX
PT New mutant Salmonella species
XX
PS Disclosure; Fig 1; 74pp; English.
XX
CC Mutant Salmonella species comprising a genetically modified msbB
CC gene can be used in the treatment of solid tumours by inducing the
CC expression of tumour necrosis factor (TNF) alpha. The mutations of
CC the Salmonella species reduce both the virulence of and septic
CC shock caused by the organism. They also increase sensitivity to
CC agents which promote eradication of the bacteria, e.g. chelating
CC agents, which facilitates eradication of the bacteria from the body
CC after treatment. The modified Salmonella induce TNF-alpha less than
CC the wild type bacteria and have reduced ability to directly kill
CC normal mammalian cells or cause systemic disease compared to the
CC wild type strain. The mutant Salmonella species can be used for
CC inhibiting growth or reducing the volume of a solid tumor cancer,
CC e.g. melanoma, colon carcinoma, lung cancer, liver cancer,
CC kidney cancer, prostate cancer and breast cancer.
XX
SQ Sequence 323 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 127.50 | Length: | 290 |
| Ratio: | 0.885 | Gaps: | 10 |
| Percent Similarity: | 49.655 | Percent Identity: | 24.138 |

alignment_block:
US-09-303-518D-569 x AAW85665 ..

Align seg 1/1 to: AAW85665 from: 1 to: 323

```
 64 GCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCT 113
    |||   :::      ::::::|||   |||   |||            ||
31 AlaAlaMetAlaGlyIleAlaLeuThrProAlaSerPheArgAspProLe 47

114 GGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCG 163
    |       ||||||:::::::|||      |||   |||:::::::|||
 47 uLeuAlaThrLeuGlyArgPheAlaGlyArgLeuGlyLysSerSerArgA 64

164 CGCGCATCGTCGCCAATATGCGT.............CAGGCAGGCATGAAT 201
      |||   :::   |||:::                |||   :::
 64 rgArgAlaLeuIleAsnLeuSerLeuCysPheProGlnArgSerGluAla 80

202 CCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGG 251
                  |||      :::|||||||   :::   ::::::
 81 GluArgGluAlaIleValAspGluMetPheAlaThrAlaProGlnAlaMe 97

252 TTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAA 301
          :::|||          :::|||   |||:::   |||:::
 97 tAlaMetMetAlaGluLeuAlaMetArgGlyProLysLysIleGlnGlnA 114

302 TGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAA 351
                ::::|||   |||   :::::::::         :::
114 rgValAsp...TrpGluGlyLeuGluIleIleGluGluMetArgArgAsn 129
```

```
 352 CACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGG 401
     |||   :::::::|||:::::::|||||||      :::   |||:::
 130 AspGluLysValIlePheLeuValProHisGlyTrpGlyValAspIlePr 146

 402 CGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACA 451
      :::      ::::::::::|||           :::::::|||||:::
 146 oAlaMetLeuMetAlaSerGln...GlyGlnLysMetAlaAlaMetPheH 162

 452 AACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTT 501
                      ::::::|||   |||   ::::::   |||
 162 isAsnGlnGlyAsnProValPheAspTyrIleTrpAsnThrValArgArg 178

 502 CGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAAT 551
     |||      |||:::             :::   :::|||:::|||   ::
 179 ArgPheGlyGlyArgLeuHisAlaArgAsn...AspGlyIleLysProPh 194

 552 CATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCC......G 595
     :|||::::::::::|||:::|||                 ||||||     |
 194 eIleGlnSerValArgGlnGlyTyrTrpGlyTyrTyrLeuProAspGlnA 211

 596 ACCACGTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTC 645
     |||||   |||:::   :::|||               :::||||||||||
 211 spHisGlyProGluHisSerGlu.............PheValAspPhe 222

 646 TTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGT 695
     |||:::           |||:::         ::::::|||      ||
 223 PheAlaThrTyrLysAlaThrLeuProAlaIleGlyArgLeuMetLysVa 239

 696 CAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTG..... 735
     |                    |||      |||:::
 239 l.....................CysArgAlaArgValIleProLeuP 248

 736 ..CCTGGCGGACAAGGT............TTCGATTTGCACATCCGCCCC 771
       |||      :::|||            :::   ::::::|||||||||
 248 heProValTyrAsnGlyLysThrHisArgLeuThrIleGlnIleArgPro 264

 772 GTCCAAGGGGAA...TTGAACGGCGACAAAGCCCATGATGCCGCCGTGTT 818
       :::   :::   |||::::::|||           |||      ::
 265 ProMetAspAspLeuLeuThrAlaAspAspHisThrIleAlaArgArgMe 281

 819 CAACCGCAATGCCGAATATTGGATACGCCGTTTTCCGACGCAGTATCTGT 868
     :|||::::::::|||   ::::::         |||   ||||||   :
 281 tAsnGluGluValGluIlePheValGlyProHisProGluGlnTyrThrT 298

 869 TTATGTACAACCGCTACAAA 888
     :::::   :::      |||
 298 rpIleLeuLysLeuLeuLys 304
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG10614

seq_documentation_block:

ID ABG10614 standard; Protein; 437 AA.
XX
AC ABG10614;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #10605.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS74801.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 40973; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 437 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 122.00 | Length: | 305 |
| Ratio: | 0.904 | Gaps: | 19 |
| Percent Similarity: | 44.262 | Percent Identity: | 27.213 |

alignment_block:
US-09-303-518D-569/rev x ABG10614 ..

Align seg 1/1 to: ABG10614 from: 1 to: 437

```
 842 ATCCAATATTCGGCATTGCGGTTGAACACGGCGGCATCATGGGCT..... 798
     :::|||   |||:::   |||   :::::::::   |||||||||
 142 ValGlnCysSerSerHisArgAlaArgValValLeuSerTrpAlaAspTy 158

 797 .....TTGTCGCCGT 788
                 |||   |
 158 rLeuArgArgValAlaProThrAlaAlaLeuArgGluValAlaSerLeuP 175

 787 TCAATTCCCCTTGGACGGGGCGGATGTGCAAATCGAAACCTTGTCCGCCA 738
     ||:::::::|||   :::   :::|||               |||||||||
 175 heArgLysProSerAlaGluGluMetLeu.............CysProPro 187

 737 GGCAGGCGT.....TCGCAG..... 723
     :::|||:::      ||||||
 188 AlaArgGlnArgGlyGluAlaArgGluLeuSerGlnAsnProIleIleAr 204

 722 ..................CAGAAAAACAGGGTTTTCACGCCTTTGACGT 692
                        |||   :::|||   :::   :::   |
 204 gAlaGlyGluGlyArgLeuGlnGlyHisArgGlnValAlaThrValIleC 221

 691 GTGCCAATTTTGCCGCCAGCGTCATGGTATAGGCAGGTTTGCCGAAGAAA 642
     |||||           |||||||||         ||||||
 221 ysAlaProAlaProPheSerValMet.........GlyLeu......... 231
```

```
641 TCCACCCATACGCCTTCCCCGCCTTCTTGAGGGGAGGGGACGTGGTCGGG 592
232 .........SerProGluLeuGlnGlnCysIleValGlyAsnPheAlaSe 245

591 CAGG...ACGATGGTTGCTTCGCCCGAACGCAGGGCTTTGATGATTTGTT 545
245 rArgTyrThrMetThrLysSer.........SerAlaValLeuPheIleL 259

544 TGACCCCTTGTATGCTGGTAGGCGCGGTTTTTCCTTTGCCGCGAACCCTG 495
259 eu............IlePheSerLeuIlePheLysLeu...GluGluLeu 270

494 CCCGCCTGCATGATTTTGTCTATCGCTTTGATTTTCGGCGGTTTGTACAT 445
271 ArgAlaAlaLeuValLeuValValLeuLeuIleAlaGlyGlyLeuPheMe 287

444 G.......................GCGGTCAGCGGGAACGGAAGCTGCT 419
287 tPheThrTyrLysSerThrGlnPheAsnValGluGlyPheAlaTrpCysT 304

418 GGCTGATGTAGCGTCCGCCCAAATCGTAGCTG.................. 387
304 rp............GlyProArgSerSerValAlaPheAlaGlyProSer 316

386 CCGATGTGCGGCGTG......ATGAATAGCAGCCCTTCGTGTTTGTCCAA 343
317 ProArgCysSerCysArgArgLeuAsnSerAlaSerArgIleProSerTh 333

342 AGCCTGCTGCACATGTTCCCAGCCGTGTACCGCTTTGAACATTGTTTCTA 293
333 rProCysSerThrCysSerHisSerCys..................... 342

292 TGTCTTCCGGTTTTCTGAAAAACGCGGGGGCAAGTTCCAAACCGCCTTTT 243
343 ..SerTrpGlyLeuPhe.....................ProLeuPhe 350

242 GCCGTTTCCGCAAAAACGGCT......TTGACCGTTTTGGGGTCGGGATT 199
351 AlaValPheGluGlyThrGlyLeuLeuLeuArgValLeuGlySerLeuPh 367

198 CATGCCTGCCTGACGCATATTGGCGACGATGCGCGCGCGGTCTTCCTTTA 149
367 eLeuGlyGly......IleLeuAlaPheGlyLeuGlyPheSerGluPheL 382

148 AAAGGTAAAACGCCAGATGTCCGAGCCGG..................... 120
382 euLeuValSerSerArgThrSerSerLeuThrLeuSerIleAlaGlyIle 398

119 TTTCCCAGCGTGTGC 105
399 PheLysGluValCys 403
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY51614

seq_documentation_block:

```
ID   AAY51614 standard; Protein; 306 AA.
XX
AC   AAY51614;
XX
DT   26-MAY-2000  (first entry)
XX
DE   S. avermitilis HPPD protein.
XX
KW   DOXS; 1-deoxy-D-xylulose-5-phosphate synthase; HPPD; GGPPOR; plant;
KW   p-hydroxyphenylpyruvate dioxygenase; tocopherol; vitamin K; chlorophyll;
KW   geranylgeranyl-pyrophosphate oxidoreductase; carotenoid; transgenic.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200008169-A1.
XX
PD   17-FEB-2000.
XX
PF   30-JUL-1999;   99WO-EP05467.
XX
PR   05-AUG-1998;   98DE-1035219.
PR   01-OCT-1998;   98DE-1045216.
PR   01-OCT-1998;   98DE-1045224.
PR   01-OCT-1998;   98DE-1045231.
XX
PA   (SUNG-) SUNGENE GMBH & CO KGAA.
XX
PI   Reindl A,  Mejia PL,  Palmas JME,  Gracia MAC,  Ebneth M,  Herbers K;
XX
DR   WPI; 2000-195578/17.
DR   N-PSDB; AAZ88978.
XX
PT   Use of DNA encoding 1-deoxy-D-xylulose-5-phosphate synthase to produce
PT   plants with increased tocopherol, vitamin K, chlorophyll and carotenoid
PT   content  -
XX
PS   Example 10; Page 80-82; 94pp; German.
XX
CC   This invention describes the novel use of a DNA sequence encoding
CC   1-deoxy-D-xylulose-5-phosphate synthase (DOXS), and optionally
CC   p-hydroxyphenylpyruvate dioxygenase (HPPD) and/or geranylgeranyl-
CC   pyrophosphate oxidoreductase (GGPPOR), to produce a plant with increased
CC   tocopherol, vitamin K, chlorophyll and/or carotenoid content. Transgenic
CC   plants containing DOXS DNA coding sequences can be used for production of
CC   plant and bacterial DOXS which also have increased tocopherol, vitamin K,
CC   chlorophyll and/or carotenoid content. The test system can be used to
CC   identify inhibitors of DOXS. This sequence represents the Streptomyces
CC   avermitilis HPPD protein described in the method of the invention.
XX
SQ   Sequence   306 AA;
```

alignment_scores:

```
            Quality:  121.50          Length:    324
              Ratio:  0.988             Gaps:     16
Percent Similarity:  37.963   Percent Identity: 24.383
```

alignment_block:

US-09-303-518D-569 x AAY51614 ..

Align seg 1/1 to: AAY51614 from: 1 to: 306

```
 36 GCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCC 85
 34 AlaAspHisThrProHisSerArgHisArgProAlaGlyArgProLeuPr 50

 86 TGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTG 135
 50 oGlyGluGly..................................... 53

136 GCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATATGCG 185
 54 .............AsnGlyArgGlyArgLeuArgArgArgGlnArgGln 65

186 TCAGGC..................................................... 191
 66 AlaGlyArgAlaLeuLeuHisArgLeuArgHisAlaAlaCysGlyValLe 82

191 ..................................................... 191
 82 uArgThrGlyGluArgGlnProArgAspArgPheValArgProHisGlnA 99

192 .......................AGGCATGAATCCCGACCCCAAAAC 215
 99 rgLeuGlyThrLeuArgProHisLeuArgHisGlnAlaArgHisProLeu 115

216 GGT......CAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGAAC 259
116 GlyProLeuProArgArgProCysGlyArgAlaArgArgArg........ 129

260 TTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGTTCAAA 309
```

```
                    ::::::      |||||||||
130 ..............ArgArgArgProArgHisArg..... 136

310 GCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGG 359
                                |||  |||:::  ||||||
137 ........................GlyProGlyArgProArgArg 143

360 GCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCT 409
        ::::::     |||:::|||
144 ProArgValArgAspArgAlaArgArg..................... 152

410 ACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCG 459
                           |||     |||::::::     ||||
153 ..................ProLeuGlyArgArgAlaValArgAlaG 162

460 AAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAA 509
    ||:::         |||   :::|||                :::|||
162 luGly.........ArgAlaArgHis.............GlyArgPro 170

510 AGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAG 559
    |||::::::|||        |||  |||   :::     |||
171 ArgArgAspArg.........HisLeuArgGlnAspProProHis..... 182

560 CCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCA.......... 599
                           |||   ||||||
183 ........................ProArgArgProAspArgLeuA 190

600 ..CGTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTT 647
      ||||||||||||   |||   ||||||||||||
190 rgArgProLeuProProArgLeuArgGlyArgArgProAsp......... 203

648 CGGCAAACCTGCCTA...TACCATGACGCTGGCGGCAAAATTGGCACACG 694
    |||:::|||:::        |||     |||   :::         ||
204 ArgArgThrAlaArgProProHisLeuProGlyHisArgProLeuArgAr 220

695 TCAAAGGCGTGAAAACCCTGTTTTTCTGCTG...CGAACGCCTGCCTGGC 741
    ||||||||||   :::|||         :::   |||    |||
220 gGlnArgArgAlaArgProAspGluArgMetGlyArgLeu.LeuGlnGln 236

742 GGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGG 791
    |||:::|||:::     |||   :::   |||:::|||:::   :::
237 GlyHisGlyLeuHisGluHis...GluGlyValArgGlyArgArgHisAr 252

792 CGACAAAGCCCATGATGCCGCCGTGTTCAACCGCAATGCCGAATATTGGA 841
     |||::::::      |||   |||   :::|||:::   ::::::
252 gAspArgValLeuGlyAlaAspValGluGlyArgGlyArgArgHisAlaG 269

842 TACGCCGTTTTCCGACGCAG 861
        :::   |||   |||
269 lnGlyGlnValProAspGln 275
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39201

seq_documentation_block:

ID AAM39201 standard; Protein; 565 AA.
XX
AC AAM39201;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2346.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58357.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2346; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 565 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 121.00 | Length: | 386 |
| Ratio: | 0.818 | Gaps: | 17 |
| Percent Similarity: | 38.342 | Percent Identity: | 22.798 |

alignment_block:
US-09-303-518D-569 x AAM39201 ..

Align seg 1/1 to: AAM39201 from: 1 to: 565

```
 18 CAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCC 67
    :::|||     |||   |||:::         ||||||||||:::|||
 97 GluAlaGlyLeuProArgAlaArgAlaProHisHisProValGlnArgAl 113

 68 TGCTCAA.....................ATGCCTCTCCCT.......... 86
     ::::::                    :::|||||||||
113 aGlyHisHisHisAlaAspGluLeuThrValProLeuProGlnGlyGlyG 130

 87 ..........................................GCTGCCGCT 95
                                                ::::::
130 lyGlnArgGlyAlaGlyValArgArgGlyLeuArgGlnGlyProGlyGly 146

 96 TTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACC 145
    |||   :::|||         :::|||||||||
147 PheThrGlnAla.........GluProAlaArg................. 154
```

```
146 TTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAATAT..... 182
155 ....ArgGlyArgGlyArgCysArgArgArgArgArgArgGlyAlaL 170

183 .....GCGTCA 188
170 euGlyHisGlyHisGlyLeuArgAlaProArgArgAlaGlyProGlyGly 186

189 GGCAGGCATGAATCC.....CGACCCCAAAACGGTCAAAGCCG 226
187 GlyArgAlaArgSerAlaCysLeuArgGluProGluGlnLeuArgGlyAr 203

227 TTTTTGCGGAAACGGCAAAAGGCGGTTTGGA.....ACTT 261
203 gArgArgGlyArgGlyGlyArgArgAspGlyArgLeuArgCysArgThrC 220

262 GCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGTTCAAAGC 311
220 ysAlaLeuProAlaAlaAlaThrProArg..... 229

312 GGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA..... 350
230 .....ArgValGlyAlaAlaAlaProGlyCysAlaGlnGlyArgAlaTh 244

351 .....ACACGAAGGGCTGCTATTCATCAC.....GCCGCAC 381
244 rAspGlyAlaArgArgAlaGluLeuArgArgGluProAlaValValAlaH 261

382 ATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCTTCCGTT 431
261 isArgHisGlyHisAlaGly.....AlaHisGlnGlyGlyAlaGlnAla 275

432 CCCGCTGACCGCCAT..... 446
276 AlaAlaGlnProHisArgArgLeuGlnValProGlnAlaGlnAlaGlyAl 292

447 .....GTACAAACCGCCGAAAATCAAAGCGATAGACAAAATC 483
292 aHisLeuAlaProGlyArgGluSerGluAspProGlnGluSerGluHisG 309

484 ATGCAGGCGGGCAGGGTTCG...CGGCAAAGGAAAAACCGCGCCTACCAG 530
309 lyAlaGlyValHisGlyGluProAlaAlaArgAlaGlyGlyAlaAlaGln 325

531 CATACAAGGGGTCAAACAAATCATCAA..... 557
326 AlaGluSerProGlnProArgGlnGlnArgLeuProAlaAlaAlaProAl 342

558 .....AGCCCTGCGTTCGGGCGAAGCAACCATCGTC 588
342 aProGlyAlaArgValLeuSerProArgAlaGlyArgMetArgGlyHisP 359

589 CTGCCCGA..... 596
359 roProGlnGlyAlaGlySerArgGlyGlyValValGlyAlaProAspLeu 375

597 .....CCACGTCCCCTCCCCTCAAG 616
376 GluArgValArgProTrpGlyProProProProArgValCysProGlyTh 392

617 AAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATACC 666
392 rGlnArgGlyArgGlyProArgGlyPheProProGluGlyAla.....G 407

667 ATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTT 716
407 lnAspSerGluGlyAlaProArgThrArgGln..... 417

717 TTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCC 766
418 .....AlaGlyPr 420

767 GCCCCGTCCAAGGGGAATTGAACGGCGACAAAGCCCATGATGCCGCCGTG 816
```

```
420 oProAlaProGlyGlyGlyGluArgMetThrProProProSerArgCysL 437

817 TTCAACCG 824
437 euPhePro 439
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365

seq_documentation_block:

ID ABG20365 standard; Protein; 1098 AA.
XX
AC ABG20365;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20356.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS84552.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 50724; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1098 AA;

alignment_scores:
Quality: 121.00 Length: 382

```
            Ratio:  0.776                Gaps:     17
Percent Similarity:  40.838   Percent Identity: 24.346

alignment_block:
US-09-303-518D-569 x ABG20365   ..

Align seg 1/1  to: ABG20365  from: 1  to: 1098

     45 CATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGC 94
        |||      |||:::||||||   :::      |||   |||   |||||
    525 HisProProProAlaAspArgArgSerProProProArgProProAlaAl 541

     95 TTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTAC 144
        |                 |||:::   :::::::::   |||:::
    541 aArgArgArgArgProProGlyGlnAspSerGlnAlaProGlyAla.... 556

    145 CTTTTAAAGGAAGACCGCGCGCG...CATCGTCGCCAATATGCGTCAGGC 191
                |||||||||||||      |||      :::      |||
    557 ........ArgArgProArgAlaThrAspArgSerProHisProArgGly 570

    192 AGGCATGAA.....................................TC 202
        |||   :::                                     ::
    571 ArgThrGlnThrHisAlaProThrAspAlaAlaHisArgThrProProTh 587

    203 CCGACCCCAAAACGGTCAA.........................AGCC 225
        :|||||||||||:::   :::                      :::|
    587 rArgProGlnArgProArgArgArgArgGluArgLysGluThrAspGluA 604

    226 GTTTTTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTT 275
        ||         :::|||||||:::::            ||||||
    604 rgLysGlnArgArgGlyLysGluLysGlyThrLysGlnProArgLysArg 620

    276 CAGAAAACCGGAAGACATAGA............................ 296
        ||||||      |||   |||
    621 GlnLysGlyThrArgLysArgArgArgGluAsnGluLysLysArgGluAr 637

    296 ............................................... 296

    637 gArgGluLysArgArgGlnGluArgLysLysLysLysLysArgLysG 654

    297 ..........................AACAATGTTCAAAGCGGTACA 317
                                   ::::::   ::::::
    654 luGluAlaArgArgArgArgLysGluGlnLysLysGluGluLysArgArg 670

    318 CGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTAT 367
        |||      |||      |||      ::::::|||:::|||
    671 ArgSerGluThrLysArgAlaLysHisAlaLysThrLysArg........ 684

    368 TCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGC 417
           :::      :::|||:::   |||   |||:::|||   :::
    685 ....ArgGluGluArgArgGluArgArgArgGlyAsnThrArgArg.... 698

    418 CAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAA 467
                           |||   :::::::::::::|||:::::::
    699 ......................ArgThrThrArgAlaGlyGluGluArg 707

    468 AGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAA 517
        :::   |||:::::::::|||   |||:::   ::::::   |||||||::
    708 GlyArgArgArgGluArgAlaArgGlyArgGluAlaGlnThrArgLysAr 724

    518 CCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGT 567
        ::::|||      |||||||:::   :::::::::::::::::|||
    724 gHisAla......HisThrGlnThrLysAsnLysGluArgThrPro.... 737

    568 TCGGGCGAAGCAACCATCGTCCTGCCCGACCACGT.........CCCCTC 608
            |||||||||   |||   ||||||   |||          |||
    738 .....ArgSerAsnLysArgSerAlaArgAlaArgLysAsnThrProAla 752

    609 CCCTCAAGAAGGCGGGGAAGGCGTATG..............GGTGGATT 643
           :::|||||      ||||||                  :::|||
    753 ThrAlaArgArgProArgArgArgArgAlaArgThrProProAlaGlyAl 769

    644 TCTTCGGCAAACC..................................... 656

           :::|||:::
    769 aProGlnGlnAlaGlnAsnAlaProProAlaProAlaThrAlaArgAsnA 786

    657 .....................................TGCCTATAC 665

    786 spLysGluLysThrProArgProProAlaProAlaThrArgThrThrArg 802

    666 CAT.....................................GACGCTGGCGG 679
        |||                                    ||||||   ||
    803 HisProGluAspProAlaProProProAlaHisProGlyAspAlaArgGl 819

    680 CAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAA 729
        |            |||:::||||||::::::|||:::   |||   |||
    819 yProAlaArgProArgLysArgArgArgArgProAla...LeuArgArgH 835

    730 CGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGT...CCA 776
           |||   :::   |||   |||   :::   |||||||||   |||
    835 isAspAlaGlnHisProArgGlyArgProGlyThrProProArgHisPro 851

    777 AGGGGAATTGAACG....GCGACAAAGCCCATGATGCCGCCGTGTTCA.. 820
        |||      |||      ||||||   |||         |||:::|||
    852 ArgProProGluProProAlaThrProProArgAlaHisProAlaSerGl 868

    821 ....ACCGCAATGCCGAATATTGGATACGCCGTTTTCCGACGCAGT 862
              |||   |||:::         |||:::   |||||||||
    868 uArgProAlaProProHisProProLysAlaAlaCysArgArgSer 883

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21221

seq_documentation_block:
ID   ABG21221 standard; Protein; 1243 AA.
XX
AC   ABG21221;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #21212.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS85408.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 51580; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
```

```
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1243 AA;


alignment_scores:
           Quality:  120.50              Length:    216
             Ratio:  1.181                Gaps:      9
 Percent Similarity:  47.222   Percent Identity: 24.537

alignment_block:
 US-09-303-518D-569 x ABG21221   ..

 Align seg 1/1  to: ABG21221  from: 1  to: 1243

      66 CCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGG 115
         ||||||      |||   ||||||||||||                 |||||
     417 ProAlaProAlaProAlaProAlaAlaAla..............AlaGl 428

     116 GAAACCGGCTCGG.................................... 128
         |   |||:::|||
     428 yProProThrArgLysLysSerArgArgGluArgLysArgSerGlyGluA 445

     129 ....................ACATCTGGCGTTTTACCTTTTAAAGGAAGA 158
                              :::||||||:::   |||
     445 laLysGluAlaAlaSerSerSerSerGlyThrGlnPro.........Ala 458

     159 CCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAATCCCGACC 208
         |||   |||            :::   |||   :::|||:::||||||
     459 ProProAlaProAlaSerProTrpAspSerLysLysHisArgSerArgAs 475

     209 CCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGAA 258
          ::::::      |||:::      :::::::::::::||||||
     475 pArgLysProGlySerHisAlaSerSerSerAlaArgArgArg....... 489

     259 CTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGTTCAA 308
                                 :::|||   |||::::::   :::
     490 ........................SerArgSerArgSerArgSerArg 497

     309 AGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAG 358
         |||      |||   :::      :::            :::   |||||
     498 SerThrArgArgArgSerArgSerThr............AspArgArgAr 510

     359 GGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGC 408
         |::::::   :::::::::      |||:::   |||      |||
     510 gGlyGlySerArgArgSerArgSerArgGluLysArg......ArgArgA 525

     409 TACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCC 458
           ::::::   |||:::      ||||||              :::::::::
     525 rgArgArgSerAlaSerProProProAlaThrSerSerSerSerSerSer 541

     459 GAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCA 508
         :::::::::      |||   :::|||      |||   ||||||:::::
     542 ArgArgGluArgHisArgGlyLysHisArgAspGlyGlyGlySerLysLy 558

     509 AAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAA 558
         ::::|||:::|||:::      ::::::||||||:::   :::
     558 sLysLysLysArgSer......ArgSerArgGlyGluLysArgSer.... 571

     559 GCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCTC 608
```

```
                  |||   ::::::::::::         |||   |||
     572 ..........GlyAspGlySerGluLysAlaProAlaProAlaProPro 584

     609 CCC..............................TCAAGAAGGCGGGGA 626
         |||                             ||||||||||||||||
     585 ProSerGlySerThrSerCysGlyAspArgAspSerArgArgArgGly 600

 seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

 seq_documentation_block:
 ID   AAB59817 standard; Protein; 999 AA.
 XX
 AC   AAB59817;
 XX
 DT   04-APR-2001  (first entry)
 XX
 DE   TutD protein #8.
 XX
 KW   Toluene degradation; enzyme; waste degradation; TutD.
 XX
 OS   Thauera aromatica.
 OS   Xanthomonas maltophilia.
 OS   Geobacter metallireducens.
 OS   Azoarcus tolulyticus.
 XX
 PN   WO200072650-A2.
 XX
 PD   07-DEC-2000.
 XX
 PF   24-MAY-2000; 2000WO-US14298.
 XX
 PR   01-JUN-1999;   99US-0323872.
 XX
 PA   (UYOH-) UNIV OHIO.
 XX
 PI   Coschigano PW;
 XX
 DR   WPI; 2001-041080/05.
 DR   N-PSDB; AAF23625, AAF23627.
 XX
 PT   Composition comprising toluene degrading enzyme useful for biological
 PT   treatment of organic compounds, especially for degrading toluene or its
 PT   analogs -
 XX
 PS   Disclosure; Fig 5; 122pp; English.
 XX
 CC   The present invention relates to toluene degrading enzyme genes and
 CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
 CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
 CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
 CC   toluene degrading enzymes are useful for biological treatment of organic
 CC   compounds and in particular for the degradation of toluene and its
 CC   analogs contained in liquid or solid waste source. The present sequence
 CC   is a protein sequence for toluene degrading enzyme, TutD.
 XX
 SQ   Sequence   999 AA;


alignment_scores:
           Quality:  120.00              Length:    407
             Ratio:  0.732                Gaps:     27
 Percent Similarity:  40.295   Percent Identity: 25.061

alignment_block:
 US-09-303-518D-569 x AAB59817   ..

 Align seg 1/1  to: AAB59817  from: 1  to: 999

      29 CCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCC.......... 67
         |||||||||          :::|||||||||      |||
     212 ProLeuCysGlyAlaThrAlaThrSerCys......ProArgArgArgAr 226

      68 .TGCTCA...AATGCCTCTCCCTGCTGC.................CGCT 95
```

```
     |||||||   :::|||||||      |||                        |||
226 gCysSerIleGlyAlaSerSerGlyCysProHisProProValArgArgS 243

 96 TTCCTGTC.........................TGCACA......... 109
    ||||||                           ||||||
243 erProValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArg 259

110 .........CGCTGGGAAACCGGCTCGGAC..............ATCTG 135
             |||      |||:::   |||                 ||
260 GlyArgPheArgGlyProThrSerArgAspThrGlyArgArgArgCysTr 276

136 GCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGC......... 176
    ||||::: |||         |||   |||:::   ||||||
276 pArgTrp.ProArgProArgArgCysArgCysSerArgArgTrpGlyArg 292

177 ...CAATATGCGTCAGGC.................AGGCATGAATCCCG 205
          :::||||||||||                 |||:::   |||::
293 ProLeuTrpAlaSerGlyCysProArgAlaArgTrpArgArgGlySerAs 309

206 ACCCCAAAACGGTCAAAGC.................CGTTTTTGCGGAA 237
    :   ::::::|||:::|||                 |||   ||||||:
309 nTrpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCysGlyA 326

238 ACGGCAAAAGG..............CGGTTTGGAACTTGCCCC...... 266
    ::      |||              |||   :::   ||||||
326 rgArgValArgSerAspThrSerAlaArgArgSerArgCysProAlaSer 342

267 CGCGTTTTTCAGAAAACCGGAAGACATAGAAAC................ 299
          ::::::   |||||||||   |||:::
343 SerProIleArgTrpThrGlyArgCysArgArgTrpArgArgProLeuGl 359

300 ..........AATGTTCAAAGCGGTACACGGCTGGGAACA......TGTG 333
            ::::::          :::|||   |||         |||:
359 yCysSerProArgAlaThrCysThrAlaArgCysGlyArgAspGlyCysS 376

334 CAGCAGGCTTTGGACAA.............................. 350
    ::|||   |||||:::
376 erAlaPhePheGlyAsnProLeuHisArgSerLeuArgGlyProTrpAla 392

351 ...............................ACACGAAGGGCTGCTAT 367
                                  |||||||||:::|||::
393 AlaProPheArgAlaHisArgSerArgSerThrThrArgArgCysAlaVa 409

368 TCATCACGCCGCA...CATCGGCAGCTACGATTTGGGCGGACGCTACATC 414
    ::::   ::::::   |||   :::          |||      |||:
409 lArgGlySerSerArgHisAspArgThrAlaSerThrArgArgProHisL 426

415 AGCCAGCAGCTTCCGTTCCCGCTGACCGCCAT................. 446
    ::|||      :::      :::|||   |||
426 ysProProLysGlyCysAlaThrAspIleHisSerGlyArgTyrCysTrp 442

447 GTACAAACCGCCGAAAATCAAAGCGAT..................... 473
       :::||||||::::::::::::
443 ProArgThrAlaSerSerArgAlaAlaSerGlyAlaSerAlaLysArgTh 459

474 .......AGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAA 516
          |||::::::   :::          ::::::   |||:::|||   :
459 rArgLeuArgArgArgSerCysProValArgSerProArgArgArgGlyT 476

517 ACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCG 566
    ::|||||      |||:::         ::::::   :::   |||:::
476 hrArgAlaAlaTrpHisSerAlaCysGlySerSerSerArgArgProSer 492

567 TTCGGGCGAAGCAACCATCGTCCTGCCCGACCA................ 599
       ||||||           |||   ||||||
493 SerGlyArgProTrpSerValProIleArgProSerSerIleCysGlyAr 509

600 ..................................CGTCCCCTC.... 608
                                     |||||||:::
509 gAlaValGlyLeuThrSerProSerSerProLeuAsnArgProPheAlaA 526

609 ......................CCCTCAAGAAGGCGGGGAAGGCGTATG 635
                         |||   ||||||:::::|||||||
```

```
526 rgArgSerAlaProAlaSerThrProCysArgArgHisAsnArgArgArg 542

636 GGTGGA............TTTCTTCGGCAA...ACCTGCCTATAC..... 665
       |||            |||   |||:::   :::|||   :::
543 TyrGlySerArgArgProPheArgArgArgPheAlaCysSerTrpSerSe 559

666 ....CATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACC 711
        ||||||   :::::::::            ||||||         |
559 rGlnHisAspProAlaSerGln......AspProGlnArgGlyThrCysP 574

712 CTGTTTTTCTGCTGCGAACG......CCTGCCTGGCGGACAAGGTTTCGA 755
    ||          ||||||:::      |||:::|||      |||
574 ro.........LeuArgAsnAlaCysProGlyTrpAlaProArgAlaSer 587

756 TTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGCCCATG 805
          |||   |||      |||      :::::::|||
588 ArgProHisLeuProLeuArgArgArgCysLysGluArgCysProPheAr 604

806 ATGCCGCCGTGTTCAACCG 824
    |||      ::::::|||
604 gCysSerProAlaSerPro 610
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:

ID AAB59827 standard; Protein; 1592 AA.
XX
AC AAB59827;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #4 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1592 AA;

```
alignment_scores:
            Quality:  120.00                Length:    407
              Ratio:  0.732                  Gaps:     27
 Percent Similarity:  40.295   Percent Identity: 25.061

alignment_block:
 US-09-303-518D-569 x AAB59827   ..

 Align seg 1/1  to: AAB59827  from: 1  to: 1592

      29 CCCCTTTGCGAACCGCCATGCACATCCTGTTGACCGCCC..... 67
         |||||||||           :::||||||||||      |||
     805 ProLeuCysGlyAlaThrAlaThrSerCys......ProArgArgArgAr 819

      68 .TGCTCA...AATGCCTCTCCCTGCTGC.................CGCT 95
          ||||||   :::||||||      |||                  |||
     819 gCysSerIleGlyAlaSerSerGlyCysProHisProProValArgArgS 836

      96 TTCCTGTC.........................TGCACA......... 109
           ||||||                         ||||||
     836 erProValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArg 852

     110 .........CGCTGGGAAACCGGCTCGGAC..............ATCTG 135
                 |||      |||:::   |||                 ||
     853 GlyArgPheArgGlyProThrSerArgAspThrGlyArgArgArgCysTr 869

     136 GCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGC......... 176
         ||||::: |||          |||   |||:::   ||||||
     869 pArgTrp.ProArgProArgArgCysArgCysSerArgArgTrpGlyArg 885

     177 ...CAATATGCGTCAGGC..................AGGCATGAATCCCG 205
            :::||||||||||                   |||:::   |||::
     886 ProLeuTrpAlaSerGlyCysProArgAlaArgTrpArgArgGlySerAs 902

     206 ACCCCAAAACGGTCAAAGC..................CGTTTTTGCGGAA 237
         :   ::::::|||:::|||                  |||   ||||||:
     902 nTrpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCysGlyA 919

     238 ACGGCAAAAGG..............CGGTTTGGAACTTGCCCC...... 266
         ::      |||              |||   :::   ||||||
     919 rgArgValArgSerAspThrSerAlaArgArgSerArgCysProAlaSer 935

     267 CGCGTTTTTCAGAAAACCGGAAGACATAGAAAC.............. 299
             ::::::   |||||||||   |||:::
     936 SerProIleArgTrpThrGlyArgCysArgArgTrpArgArgProLeuGl 952

     300 ..........AATGTTCAAAGCGGTACACGGCTGGGAACA......TGTG 333
                  ::::::          :::|||   |||         |||:
     952 yCysSerProArgAlaThrCysThrAlaArgCysGlyArgAspGlyCysS 969

     334 CAGCAGGCTTTGGACAA.............................. 350
         ::|||   |||||||:::
     969 erAlaPhePheGlyAsnProLeuHisArgSerLeuArgGlyProTrpAla 985

     351 ............................ACACGAAGGGCTGCTAT 367
                                     |||||||||:::|||::
     986 AlaProPheArgAlaHisArgSerArgSerThrThrArgArgCysAlaVa 1002

     368 TCATCACGCCGCA...CATCGGCAGCTACGATTTGGGCGGACGCTACATC 414
         ::::   ::::::   |||   :::           |||      |||:
    1002 lArgGlySerSerArgHisAspArgThrAlaSerThrArgArgProHisL 1019

     415 AGCCAGCAGCTTCCGTTCCCGCTGACCGCCAT.................. 446
         ::|||      :::      :::|||   |||
    1019 ysProProLysGlyCysAlaThrAspIleHisSerGlyArgTyrCysTrp 1035

     447 GTACAAACCGCCGAAAATCAAAGCGAT..................... 473
            :::||||||:::::::::::::
    1036 ProArgThrAlaSerSerArgAlaAlaSerGlyAlaSerAlaLysArgTh 1052

     474 .......AGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAA 516
               |||::::::   :::      ::::::   |||:::|||   :
    1052 rArgLeuArgArgArgSerCysProValArgSerProArgArgArgGlyT 1069

     517 ACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCG 566
         ::|||||      |||:::         ::::::   :::   |||:::
    1069 hrArgAlaAlaTrpHisSerAlaCysGlySerSerSerArgArgProSer 1085

     567 TTCGGGCGAAGCAACCATCGTCCTGCCCGACCA................. 599
            ||||||            |||   ||||||
    1086 SerGlyArgProTrpSerValProIleArgProSerSerIleCysGlyAr 1102

     600 ....................................CGTCCCCTC.... 608
                                            |||||||:::
    1102 gAlaValGlyLeuThrSerProSerSerProLeuAsnArgProPheAlaA 1119

     609 .....................CCCTCAAGAAGGCGGGGAAGGCGTATG 635
                              |||   ||||||:::::|||||||
    1119 rgArgSerAlaProAlaSerThrProCysArgArgHisAsnArgArgArg 1135

     636 GGTGGA............TTTCTTCGGCAA...ACCTGCCTATAC..... 665
            |||             |||   |||:::   :::|||   :::
    1136 TyrGlySerArgArgProPheArgArgArgPheAlaCysSerTrpSerSe 1152

     666 ....CATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACC 711
             ||||||   :::::::::            ||||||         |
    1152 rGlnHisAspProAlaSerGln......AspProGlnArgGlyThrCysP 1167

     712 CTGTTTTTCTGCTGCGAACG......CCTGCCTGGCGGACAAGGTTTCGA 755
         ||         ||||||:::      |||:::|||      |||
    1167 ro.........LeuArgAsnAlaCysProGlyTrpAlaProArgAlaSer 1180

     756 TTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGCCCATG 805
             |||   |||       |||       ::::::|||
    1181 ArgProHisLeuProLeuArgArgArgCysLysGluArgCysProPheAr 1197

     806 ATGCCGCCGTGTTCAACCG 824
          |||      ::::::|||
    1197 gCysSerProAlaSerPro 1203

 seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29221

 seq_documentation_block:
 ID   AAY29221 standard; Protein; 447 AA.
 XX
 AC   AAY29221;
 XX
 DT   25-OCT-1999  (first entry)
 XX
 DE   Amino acid sequence of a virulence factor encoded by ORF31266.
 XX
 KW   Human pathogen; virulence polypeptide; virulence factor;
 KW   pathogenic infection; Pseudomonas aeruginosa infection.
 XX
 OS   Pseudomonas aeruginosa.
 XX
 PN   WO9927129-A1.
 XX
 PD   03-JUN-1999.
 XX
 PF   25-NOV-1998;   98WO-US25247.
 XX
 PR   25-NOV-1997;   97US-0066517.
 XX
 PA   (GEHO ) GEN HOSPITAL CORP.
 XX
 PI   Ausubel F,  Cao H,  Drenkard E,  Goodman HM,  Mahajan-Miklos S;
 PI   Rahme LG,  Tan M,  Tsongalis J;
 XX
 DR   WPI; 1999-357851/30.
 XX
 PT   Virulence factors useful in developing disease treatments
 XX
 PS   Disclosure; Fig 3; 228pp; English.
 XX
 CC   The present sequence represents a Pseudomonas aeruginosa polypeptide
```

```
CC   sequence. P. aeruginosa is an opportunistic human pathogen present in
CC   soil water and plants. The specification describes virulence polypeptides
CC   and nucleic acid sequence encoding such polypeptides. These sequences
CC   can be used to identify a compound which is capable of decreasing the
CC   expression of a pathogenic virulence factor. Compounds that inhibit
CC   the expression or activity of virulence factor polypeptides can be
CC   used to treat pathogenic infections, especially where the infection
CC   is a P. aeruginosa infection.
CC   note: the sequences given in the specification were poorly legible, and
CC   in some instances assumptions were made as to the identity of the
CC   residue; it is therefore possible that the sequence given below is
CC   not entirely correct.
XX
SQ   Sequence   447 AA;


alignment_scores:
            Quality:  119.00              Length:    279
              Ratio:  0.944                Gaps:     14
 Percent Similarity:  45.161   Percent Identity: 26.165

alignment_block:
 US-09-303-518D-569 x AAY29221   ..

 Align seg 1/1  to: AAY29221  from: 1  to: 447

      51 CATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCT 100
         |||||||||   |||   ::::::   |||
     203 HisProValProArgArgSerArgArgPro.................Al 213

     101 GTCTGCACACGCTGGGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTA 150
         ::::::||||||||||:::   |||      :::         |||   |
     213 aAspGlyHisAlaGlyGlnThrAlaProGlyAsp.........ProGlyL 227

     151 AAGGAAGACCGCGCGCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAA 200
         ||||||||   |||                                ||||||
     227 ysGlyArgAlaArg.......................GlyArg...... 233

     201 TCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCG 250
         ||||||   :::|||:::   |||   |||::::::::::
     234 ...ArgProAlaGlyGlyArgArgArgThrCysAlaSerSerAlaProPr 249

     251 GTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACA 300
             |||   |||                          :::      |||::::
     249 oThrArgThrSerProProLysSerAlaProGlyAlaSerAlaArgThrS 266

     301 ATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAA 350
         :::::   :::::::|||      ::::::||||||           ::::::
     266 erThrThrAlaSerThrSer...SerSerCysAlaTyrThrAlaAlaArg 281

     351 ACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGG 400
         |||   |||:::      |||   |||||||::|||
     282 ThrProArgGly......HisProAlaAlaArgArgThr.......... 292

     401 GCGGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTAC 450
                       :::|||   |||      |||:::   ||||||   :
     293 .............HisProGlnAla......ProGlyArgArgHis...A 302

     451 AAACCGCCGAAAATCAAAGCGATAGACAA..................... 479
         ::   |||   :::|||:::|||||||:::
     302 rgProAlaGlyArgGlnAlaAspArgArgArgThrGlyGluAlaGluGlu 318

     480 ................................AATCATGCAGGCGGGCA 496
                                        :::||||||||||:::::
     319 LeuProLeuProGlyGlnArgProArgAlaGlyLysHisAlaGlyAlaAr 335

     497 G......GGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGG 540
         :           |||:::|||   :::   |||   :::   :::   |
     335 gLeuTyrProValArgArgArgProAspProAlaSerArgProAlaProG 352

     541 GTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCT 590
         ||:::           :::|||:::      |||:::   :::|||
     352 lyArgCysAlaGlyCysGlnProGlyArgArgArgGluProGluArgAsn 368

     591 GCCCGACCACGTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGG 640
            :::||||||   ||||||:::|||:::|||::::::         ::
     369 ArgGlnProArgGlyLeuProGlyArgHisArgAlaGlnAlaAspHisAl 385

     641 ATTTCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGC.......... 680
         :      |||         |||                 |||
     385 aGlyThrArgGlyAspProLeuGluProHisArgArgGlyProAlaProG 402

     681 ................................AAAATTGGCACACGT 695
                                        |||   |||   |||
     402 lyProAspValProLeuAspAlaLeuProProGluLysAlaGlyHisArg 418

     696 CAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGAC 745
            :::   ::::::||||||         |||         |||   ::
     419 LeuLysValLysArgProVal......ArgArgGlnAlaPheTrpPheSe 433

     746 AAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGA 782
         :   :::|||      :::|||   |||   ||||||
     433 rLeuLeuArgGlyAspGlnProGlyArg...ArgGly 444

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21540

seq_documentation_block:
ID   ABG21540 standard; Protein; 531 AA.
XX
AC   ABG21540;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #21531.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS85727.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 51899; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
```

```
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   531 AA;


alignment_scores:
            Quality:  119.00              Length:    339
              Ratio:  0.753                Gaps:     18
 Percent Similarity:  46.608   Percent Identity: 24.484

alignment_block:
 US-09-303-518D-569 x ABG21540   ..

 Align seg 1/1  to: ABG21540  from: 1  to: 531

      42 CGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCT.. 89
         ||||||   |||   :::         |||      |||   ||||||
     177 ArgHisMetHisGlnLeuProAspAlaAla***ThrProTyrProAlaTy 193

      90 ....GCCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACATCTG 135
             :::   :::|||:::|||:::::         ||| :::|||
     193 rGlySerAspTyrLeu***AlaArg***AspAlaThrAla.SerHisGln 209

     136 GCGTTTTACCTTTTAAAGGAAGACCGCGCGCGC.............ATCGT 173
         |||            :::   |||   |||                
     210 AlaSerCysThrAsnSerProAspAlaAla***ThrProTyrProAlaTy 226

     174 CGCCAATATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAG 223
          |||:::::::::::::|||      ||||||   |||:::::::::|||
     226 rAlaLysIleGluArgAla***GluAsnProValProSerAlaIleLysI 243

     224 CCGTTTTTGCGGAAA.............CGGCAAAAGGCGGTTTGGAACTT 261
         :::    |||:::              ||||||   :::   |||:::::
     243 leIleGly.ArgSerSerSerSerGlyArgGlnAlaSerProTrpGlyMe 259

     262 GCCCCCG.................CGTTTTTCAGAAAACCGGAAGACAT 293
             :                   |||::::::   ||| |||
     259 tAlaSerProHisTrpArgSerTrpArgMetGlyProAsn.GlySerGly 275

     294 AGAAACAATGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTT 343
         ::::::      :::::::::   |||::::::   |||     :::
     276 LysArgProArgArgAlaAlaGluArgIleSerCysCysArgAsnAlaGl 292

     344 TGGACAAACA......CGAAGGGCTGCTATTCATCACGCC.......... 377
         |||:::|||      :::::::|||   |||::::::
     292 nGlyArgThrHisSerGlnLysAlaTyrIleGlnGlnGlnProGlnCysC 309

     378 ........GCACATCGGCAG.................CTACGATTTGGG 401
                 :::|||   |||                 :::   :::
     309 ysValAsnValHisLeuGlnProProAspProThrProAlaGlnProSer 325

     402 CGGACGCTA.........CATCAGCCAGCAGCTTCCGTTCCCGCTGACCG 442
         :::||||||          :::||||||      :::||||||
     326 GlnThrLeuSerProProAspSerProAlaProProLeuProAlaTyrPr 342

     443 CCATGTACAA....................ACCGCCGAAAATCAAAGCG 471
             :::|||                   |||::::::   :::|||:
     342 oSerThrGlnHisProThrAlaProProProThrSerSerProMetSerS 359

     472 ATAGA..............CAAAATCATGCAGGC.........GGGCAG 497
         ::|||                :::|||:::|||          ||||||
     359 erArgIleValLysProHisProGlnHisCysGlyAlaArgProGlyGln 375

     498 GGT.................TCGCGGCAAAGGAAAAACCGCGCCTACCA 529
         |||                 :::|||:::      ::::::      ::
     376 GlyLysGluArgAspGluThrGluArgLysAspAlaLysGluArgAsnAr 392

     530 GCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCA 579
         ::::|||           |||:::::::::::::      :::   |||:::
     392 gArgThrAlaThrProThrThrArgHisThrAspArgLeuProArgThrA 409

     580 ACCATCGTCCTGCCCGACCACGTCCCCTCCCC..................... 611
         |||   |||       |||   |||   |||
     409 laHisGlyProProProProProProProProProProProProProPro 425

     612 ..................TCAAGAAGGCGGGGAAGGCGTATGGGTGGATT 643
                           |||   ||||||   ||||||
     426 ProProProProArgAspSerProArgArgProArgArg........... 438

     644 TCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACAC 693
             |||:::             :::      ||||||:::   |||
     439 ....ArgGluLysProGluGlnGlnLysGluGlyGlyGlnThrGlyArgT 454

     694 GTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGG 743
           ||||||:::|||:::                 :::      :::   |||
     454 hrGlnArgLysGluThrAlaArgArgGlyGluGluArgGluThrGlyArg 470

     744 ACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAATTGAACGG.. 791
         :::   |||   |||   ||||||||||||   :::   :::|||
     471 LysLysThrArgGlyAlaProProProArgProAlaAlaAlaArgArgPr 487

     792 ..................CGACAAAGCCCATGATGCCGCCGTGTTCAAC 822
                           ::::::   |||   :::|||:::   :::|
     487 oHisGluThrAlaGluThrLysArgArgProProAlaArgGlnHisArgP 504

     823 CGCAATGCCGA 833
         ||      |||
     504 roProArgArg 507

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW55484

seq_documentation_block:
ID   AAW55484 standard; Protein; 328 AA.
XX
AC   AAW55484;
XX
DT   30-JUN-1998  (first entry)
XX
DE   H. pylori ORF 29ae30321_34157812_f3_10 cell envelope protein.
XX
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacteria; life cycle; activator;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   bacterium.
XX
OS   Helicobacter pylori.
XX
PN   WO9737044-A1.
XX
PD   09-OCT-1997.
XX
PF   27-MAR-1997;   97WO-US05223.
XX
PR   06-DEC-1996;   96US-0761318.
PR   29-MAR-1996;   96US-0625811.
PR   02-APR-1996;   96US-0758731.
PR   25-OCT-1996;   96US-0736905.
PR   28-OCT-1996;   96US-0738859.
XX
PA   (ASTR ) ASTRA AB.
XX
PI   Alm RA,  Smith D;
XX
DR   WPI; 1997-503122/46.
DR   N-PSDB; AAV24893.
XX
PT   Helicobacter pylori nucleic acid sequences and encoded
PT   polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT   infection and for diagnosis of H. pylori infection
```

```
XX
PS   Claims 14,80; Pages 691-692; 1145pp; English.
XX
CC   This sequence is a H. pylori cell envelope protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The DNA and probes derived from it may be used for the
CC   identification of H. pylori in a sample, and the diagnosis of
CC   H. pylori infection. Nucleic acid sequences complementary to the
CC   DNA act as antisense sequences, and can be used to prevent the
CC   translation of H. pylori mRNA. Antibodies against the protein can
CC   be used in immunoassays to evaluate the abundance and distribution
CC   of H. pylori-specific antigens. The genomic sequence of H. pylori
CC   (ATCC 55679) was determined from overlapping contigs generated
CC   by mechanically shearing the bacterial DNA. The sequences were
CC   analysed for ORF of at least 180 nucleotides, and the predicted
CC   coding regions defined by computer evaluation. To identify likely
CC   H. pylori antigens for vaccine development, the amino acid
CC   sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having
CC   identified and determined the sequences of interest, particular
CC   regions can be isolated from H. pylori by PCR amplification for
CC   recombinant polypeptide production, e.g. in E. coli hosts.
XX
SQ   Sequence   328 AA;


alignment_scores:
            Quality:  118.00              Length:    237
              Ratio:  0.967                Gaps:      8
 Percent Similarity:  51.477   Percent Identity: 21.519

alignment_block:
 US-09-303-518D-569 x AAW55484   ..

 Align seg 1/1  to: AAW55484  from: 1  to: 328

     253 TTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAAT 302
         ||||||         ::::::|||         ::::::   ::::::
     106 LeuGluThrIleArgValIlePheIleProLysAspGluTyrAspAlaAr 122

     303 GTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAAC 352
          |||      ::::::      |||:::|||   ::::::|||:::|||:
     122 gPheThrLeuIleAsnGlu...GluAsnValTrpLysSerLeuAsnLysG 138

     353 ACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGC 402
         ::      :::   :::      |||:::|||   ::::::
     138 luGlyGlnAlaIleThrLeuCysMetHisPheGlyTyrTrpGluAlaVal 154

     403 GGACGCTACATCAGCCAGCAGCTTCCG...TTCCCGCTGACCGCCATGTA 449
         |||      :::::::|||          :::         ::::::
     155 GlyThrThrLeuAlaGlnTyrTyrGluAsnTyrGlyArgGlyCysLeuGl 171

     450 CAAACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGG 499
          :::      |||:::      |||:::   ::::::::::::   |||
     171 yArgLeuThrLysPheAlaProIleAsnHisMetIleMetSerArgArg. 187

     500 TTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAA 549
           ::::::   |||            ::::::|||   ::::::|||:::
     188 ..GluAlaPheGlyValArgPheValAsnLysIleGlyAlaMetLysGlu 203

     550 ATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCATC...GTCCTGCCCGA 596
         :::||||||      ::::::|||::::::   :::   :::|||   ||
     204 LeuIleLysMetTyrAsnGlnGlyAsnGlyLeuValGlyIleLeuValAs 220

     597 CCACGTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCT 646
         |:::         |||:::         :::||||||   |||   ||||
     220 pGlnAsnValValProLys.........AspGlyValValValLysPheP 234

     647 TCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACACGTC 696
         ||:::|||   |||   |||   |||:::|||:::   |||:::
     234 heAsnLysAspAlaThrThrThrThrIleAlaSerIleLeuSer..... 248

     697 AAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACA 746
                                                       ::
     249 .....................................................Ar 249

     747 AGGTTTCGATTTGCACATCCGCCCCCGTCCAAGGGGAATTGAACGGCGACA 796
         :   :::::::::   |||:::|||||      :::::|||   |||
     249 gArgTyrAsnIleAspIleGlnProValPheIleAspPheAsnAspAspT 266

     797 AAGCCCATGATGCCGCCGTGTTC.....................AAC 822
           :::|||   :::|||::::::                      :::
     266 yrSerHisTyrThrAlaThrTyrTyrProSerIleArgSerGlnIleThr 282

     823 CGCAATGCC..................................... 831
           ||||||
     283 AspAsnAlaGlnAsnAspIleLeuGluCysThrGlnAlaGlnAlaSerLe 299

     832 ....GAATATTGGATACGCCGTTTTCCGACGCAGTATCTGTTTATGTACA 877
             |||      ||||||:::   |||   :::|||:::::::::::::
     299 uCysGluGluValIleArgAsnHisProGluSerTyrPheTrpPheHisA 316

     878 ACCGCTACAAA 888
         ::|||:::|||
     316 rgArgPheLys 319

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY29184

seq_documentation_block:
ID   AAY29184 standard; Protein; 558 AA.
XX
AC   AAY29184;
XX
DT   25-OCT-1999  (first entry)
XX
DE   Amino acid sequence of a virulence factor encoded by ORF25103c.
XX
KW   Human pathogen; virulence polypeptide; virulence factor;
KW   pathogenic infection; Pseudomonas aeruginosa infection.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO9927129-A1.
XX
PD   03-JUN-1999.
XX
PF   25-NOV-1998;   98WO-US25247.
XX
PR   25-NOV-1997;   97US-0066517.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Ausubel F,  Cao H,  Drenkard E,  Goodman HM,  Mahajan-Miklos S;
PI   Rahme LG,  Tan M,  Tsongalis J;
XX
DR   WPI; 1999-357851/30.
XX
PT   Virulence factors useful in developing disease treatments
XX
PS   Disclosure; Fig 4; 228pp; English.
XX
CC   The present sequence represents a Pseudomonas aeruginosa polypeptide
CC   sequence. P. aeruginosa is an opportunistic human pathogen present in
CC   soil water and plants. The specification describes virulence polypeptides
CC   and nucleic acid sequence encoding such polypeptides. These sequences
CC   can be used to identify a compound which is capable of decreasing the
CC   expression of a pathogenic virulence factor. Compounds that inhibit
CC   the expression or activity of virulence factor polypeptides can be
CC   used to treat pathogenic infections, especially where the infection
CC   is a P. aeruginosa infection.
CC   note: the sequences given in the specification were poorly legible, and
CC   in some instances assumptions were made as to the identity of the
CC   residue; it is therefore possible that the sequence given below is
CC   not entirely correct.
```

```
XX
SQ   Sequence   558 AA;


alignment_scores:
           Quality:  117.50              Length:    371
             Ratio:  0.789                Gaps:     16
 Percent Similarity:  40.162   Percent Identity: 23.181

alignment_block:
 US-09-303-518D-569 x AAY29184   ..

 Align seg 1/1  to: AAY29184  from: 1  to: 558

      18 CAGGCTGTTTCCCCCTTTGCGAACCGCCATGCA.............CA 52
         |||||||||:::|||:::   :::|||   |||             ||
      17 GlnAlaValAlaProLeuHisHisArgSerAlaThrAlaGlyGlnGlyHi 33

      53 TCCTGTTGACCGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGT 102
         |      ||||||   ::::::::::|||   |||   :::      |||:
      33 sArgProAspArgArgGlyArgGlnProHisProHisGlyAspArgLeuG 50

     103 CTGCA.....................CACGCTGGGAAACCGGCTCGGACA 131
         ::|||                        :::      |||||||:::
      50 lnAlaGlyGlyThrGlySerSerArgProSerProAspProAlaGluAsp 66

     132 TCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGC..... 176
            :::            :::|||||||   |||:::|||::::::
      67 ArgAlaArgGlyAlaGluGluGlyArgArgArgSerHisGlnGluAlaPr 83

     177 ....CAATATGCGTCAGGCAGGCATGAATCCCGACCC............. 209
             |||   ::::::|||   ::::::::::|||
      83 oGlyGlnAlaGlyGlyGlyTyrArgGlnAlaArgAlaArgIleArgArgP 100

     209 ..................................................... 209

     100 roArgGlyAspLeuGluValArgGluGlyArgGlyAlaGlyLeuGlyAla 116

     210 ........................CAAAACGGTCAAAGCCGTTTTTGCGG 235
                                 |||   |||::::::      :::::
     117 AspProAlaGluAspArgAlaGlyGlnAlaGlyAspGlyGlyGlyAlaAl 133

     236 AAACGGCAAAAGGCGGTTT............................... 254
         ::::|||:::   |||
     133 aGlnGlyArgProArgGluHisGlyAlaHisProValProAspHisProG 150

     255 .....GGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAAAC 299
              ||||||   |||          :::   :::   :::   |||:::
     150 lyProGlyThrGlnProAlaAspGlyArgProAlaArgGlnAspArgGlu 166

     300 AATGTT......CAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTT 343
             |||      |||:::|||   |||   |||:::      :::||||||
     167 ProValAlaAlaGlnGlnGlyAspArgArgGlyAsnArgArgSerGlyPh 183

     344 T.................................GGACAAACACGAA 357
         |                                 |||::::::|||:
     183 eGlnValAspArgTyrProGlyValGluAspAlaArgGlyArgAlaArgG 200

     358 GGGCTGCTATTCATCACGCC............................. 377
         ::||||||   |||   |||
     200 luAlaAlaAlaHisGlyAlaGlyAlaAlaSerAlaSerAspArgProGly 216

     378 .................GCACATCGGCAGCTACGATTTGGGCGGACGCT 409
                           ::::::|||:::      |||:::|||
     217 ArgGlyGlySerArgArgValGlnArgArgAlaProPheAlaArg..... 231

     410 ACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCG 459
             :::|||      |||   ||||||:::|||   |||         :
     232 ....ArgProArgArgSerGluProAlaGluArgLeuValProLeuProA 247

     460 AAAATCAAAGCGATAGACAAAAT........................... 482
         ::   :::         |||:::
     247 rgProAspArgGlyGlyGlnAspArgValValGlnGlyAlaGlyArgVal 263

     483 .........CATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGC 523
                  :::   ||||||   |||:::   :::
     264 ProLeuArgTyrArgGlyGlyAlaGlyAlaAspArg.............. 275

     524 CTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGC 573
          |||   :::::::::|||:::|||            |||
     276 .TyrValArgValHisGlyGluThrLeuGlyGlyProProAspArgArgA 292

     574 GAAGCAACCATCGTCCTGCCCGACCACGTCCCCTCCCCTCAAGA...... 617
           |||:::   |||      |||   |||   ||||||:::|||
     292 laSerGlyLeuArgArgLeuArgGlyArgArgLeuProAspArgGlyAsp 308

     618 .......................AGGCGGGCAAGGCGTATGGGTGGATT 643
                                 ||||||       ||||||
     309 ProProGlnAlaLeuLeuGlyGlyAlaAlaGlyArg......GlyGlyGl 323

     644 TCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACAC 693
             :::   |||:::   |||:::      |||      |||
     323 uGlyProSerGlyCysIleGlnHisSerProProGlyAlaArgGlyArgT 340

     694 GTCAAAGGCGTGAAAACCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGG 743
               |||::::::|||::::::            |||      |||
     340 hrProAspArgGlnSerProAlaTyr...GlyGlyLeuProGlnHisArg 355

     744 ............ACAAGGTTTCGATTTGCACATCCGCCC........... 770
                        :::   ||||||:::      |||
     356 GlyGlyAspAspLeuGlnProArgPheGlyAlaAspProGlyAlaGlyAr 372

     771 .CGTCCAAGGGGA 782
          ||||||||||||
     372 gArgProArgGly 376

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB95146

seq_documentation_block:
ID   AAB95146 standard; Protein; 1197 AA.
XX
AC   AAB95146;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:17169.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17169; 2537pp + CD ROM; English.
```

```
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   1197 AA;


alignment_scores:
            Quality:  117.50              Length:    307
              Ratio:  0.864                Gaps:     17
 Percent Similarity:  44.300   Percent Identity: 24.104

alignment_block:
 US-09-303-518D-569 x AAB95146   ..

 Align seq 1/1  to: AAB95146  from: 1  to: 1197

      45 CATGCACATCCTGTTGACCGCCCT..........................  68
         ||||||:::   ::::::   |||
     879 HisAlaArgAspAlaGlnAspProSerGlnSerLeuGlyAlaProGluAl  895

      69 .GCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGA  117
          |||:::   |||   ||||||   ::::::      |||   ||||||
     895 aAlaGluArgPro...ProAlaProValLeuAspMetAlaProAlaGlyH  911

     118 AACCG.....................................------...  122
          |||
     911 isProGluGlyAspAlaGluProSerProGlyGluArgValGluAspAla  927

     123 ......GCTCGGACATCTGGCGTTTTACCTTTTAAA..............  152
               ::::::   |||      |||   |||
     928 AlaAlaProLysAlaProGlyProSerProAlaLysGluLysIleGlySe  944

     153 ..................GGAAGACCGCGCGCGCATCGTCGCCAATATG  183
                          |||:::   |||:::::::|||::::::   :
     944 rLeuArgLysValAspArgGlyHisTyrArgSerArgArgGluArgSerS  961

     184 CGTCAGGCAGG......CATGAATCCCGACCCCAAAACGGTCAAAGCCGT  227
         ::|||||:::      :::|||||||||               |||:::
     961 erSerGlyGluProAlaArgGluSerArg..............SerLys  972

     228 TTTTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCA  277
               |||:::      ||||||      ||||||||||||      ::
     973 ThrGluGlyHisArgHisArgArgArgArgThrCysProArgGluArgAs  989

     278 GAAAACCGGAAGACAT......AGAAACAATGTTCAAAGCGGTACACGGC  321
         ::::      ||||||      :::::::::       |||   ||||
     989 pArgGlnAspArgHisAlaProGluHisHisProGlyHisGlyAspArgL 1006

     322 TGGGAACATGTGCAGCAGGCTTTGGACAAACACGA.........AGGGCT  362
         ||:::                  |||:::   |||         |||:::
    1006 euSerPro..............GlyGluArgArgSerLeuGlyArgCys 1017

     363 GCTATTCATCACGCCGCACATCGG.......................... 386
         :::   ||||||:::   ||||||
    1018 SerHisHisHisSerArgHisArgSerGlyValGluLeuAspTrpValAr 1034

     387 ..................CAGCTACGATTTGGGCGGACGCTACATCAGC 417
                           :::      :::|||||             |
    1034 gHisHisTyrThrGluGlyGluArgGlyTrpGlyArgGluLysPheTyrP 1051

     418 CA..................GCAGCTTCCGTTCCCGCTGACCGCCATGTA 449
         ||                    :::          ||||||::::::
    1051 roAspArgProArgTrpAspArgCysArgTyrTyrHisAspArgTyrAla 1067

     450 CAAACCGCCGAAAAT..............CAAAGCGATAGACAAAATCA 484
               |||   :::              ::::::   |||:::::::::
    1068 LeuTyrAlaAlaArgAspTrpLysProPheHisGlyGlyArgGluHisGl 1084

     485 TGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCTACCAGCATA 534
         :   :::|||      :::|||   :::|||:::            |||:
    1084 uArgAlaGlyLeuHisGluArgProHisLysAsp............HisA 1097

     535 CAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCAT 584
         ::|||||:::   :::   :::   ||||||      |||::::::|||
    1097 snArgGlyArgArgGlyCysGlu...ProAlaArgGluArgGluArgHis 1112

     585 CGTCCTGCCCGACCACGT.......................CCCCTCCC 610
         |||||||::   ||||||                        |||   ||
    1113 ArgProSerSerProArgAlaGlyAlaProHisAlaLeuAlaProHisPr 1129

     611 CTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGC. 659
         |:::|||            |||   :::            :::::::|||
    1129 oAspArgPheSerHisAspArgThrAlaLeuValAlaGlyAspAsnCysA 1146

     660 .............CTATACCATGACGCTGGCGGCAAAATTGGCACACGT 695
                       :::::::::      :::||||||      :::|||
    1146 snLeuSerAspArgPheHisGluHisGluAsnGlyLys......SerArg 1160

     696 CAAAGGCGTGAAAACCCTGTT 716
         :::|||||||::::::   |||
    1161 LysArgArgHisAspSerVal 1167

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23029

seq_documentation_block:
ID   ABG23029 standard; Protein; 3640 AA.
XX
AC   ABG23029;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23020.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87216.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 53388; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   3640 AA;


alignment_scores:
            Quality:  117.50              Length:    261
              Ratio:  1.040                 Gaps:     13
 Percent Similarity:  43.295   Percent Identity: 26.820

alignment_block:
 US-09-303-518D-569 x ABG23029   ..

 Align seg 1/1  to: ABG23029  from: 1  to: 3640

      81 CTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGAC 130
         ||||||   |||         :::   |||:::   :::   :::
     868 LeuProArgAlaProGlyGlyGlnGlnHisGlyGluArgCysGlyGlyPr 884

     131 ATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGCGCGCATCGTCGCCAAT 180
          :::|||         :::   |||   ||||||:::   :::|||   :
     884 oGlyGlyCysAlaArgLeu***GlyAlaProArgThr...GlnArgGlyH 900

     181 ATGCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGCCGTTTT 230
         ::      ||||||      :::||||||::::::||||||   |||
     900 isGluProGlyArgGlyAlaGlyArgProArgGlyGlyGlnArgArg... 915

     231 TGCGGAAACGGCAAAAGG..............CGGTTTGGAACTTGCCC 265
            ::::::|||::::::               |||   |||
     916 GlyAlaAspGlyArgGlnHisHisValAspLeuArgHisGly........ 929

     266 CCGCGTTTTTCAGAAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTA 315
                          ||||||||||||      |||      |||
     930 ...................GlyArgHisArg...AlaValGlyProGlyG 939

     316 CACGGCTGGGAACATGTGCA..........................GCA 338
           :::   |||:::|||:::
     939 lyAsnSerGlyAlaCysValProAlaGlyThrGluLeuHisSerAspArg 955

     339 GGCTTTGGACAAACACGAAGGGCTGCTATTCATCACGCCGCACATCGGCA 388
         |||   ||||||   ||||||           :::   ||:::
     956 GlyCysGlyGlnProArgArgProProGlyProGluProAlaArgAlaGl 972

     389 GCTACGATTTGGG.................CGGACGCTACATCAGCCAG 420
          ||||||   |||                 |||   ||||||      |
     972 yLeuArgProGlyGlyAlaAlaArg***ThrArgArgLeuHisProHisA 989

     421 CAGCTTCCGTTCCCGCTGAC............................. 440
         |||||:::      |||
     989 laAla***ArgAlaAlaHisGlyLeuArgHisArgGluProGlyProLeu 1005

     441 ..............................CGCCATGTACAAACCGC 457
                                      :::|||   ::::::::
    1006 ProLeuArgLeuAspLeuArgGlyTrpLeuLeuGlnHisAspArgAlaGl 1022

     458 CGAAAATCAAAGCGATAGACAAAATCATGCA..............GGCG 492
          :      ::::::|||   |||   ||||||                |
    1022 yValProAspGlyAspThrGlnLeuHisAlaGluArgHisValProProG 1039

     493 GGCAGGGTTCGCGGCAAAGGAAAAACCGCGCC.........TACCAGCAT 533
         ||   |||:::   |||   ::::::::::::         :::|||||||
    1039 lyAlaGlyAlaValGlnProArgGluGlnGlyAlaLeuLeuHisGlnHis 1055

     534 ACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGAAGCAACCA 583
            ||||||                        |||   |||   :::::::||
    1056 LeuArgGly.....................AlaArgGlyGlyGlnArgHi 1065

     584 TCGTCCTGCCCGACCACGTCCCCTCCCCTCAAGAAGGCGGGGAAGGCGTA 633
         |      ||||||                  :::|||   ||||||
    1065 sProAlaAlaArgGluAlaValCysAlaAlaArgGlyArgGly..LeuAl 1081

     634 TGGGTGGATTTCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAA 683
          |||   :::      |||   :::|||:::|||   :::
    1081 aGlyGlyMetCysLeuAlaProValProLeuProLeuHisLeuGlyLeu. 1097

     684 ATTGGCACACGTCAAAGGCGTGAAAACCCTGT 715
           ||||||   ::::::         ||||||
    1098 ..TrpHisArgGlySerArgProHisProCys 1107

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38613

seq_documentation_block:
ID   AAY38613 standard; Protein; 318 AA.
XX
AC   AAY38613;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigenic protein encoded by ORF8.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 170-171; 524pp; English.
XX
```

```
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   318 AA;


alignment_scores:
           Quality:  116.50            Length:    298
             Ratio:  0.850              Gaps:     14
 Percent Similarity:  45.973   Percent Identity: 25.839

alignment_block:
 US-09-303-518D-569 x AAY38613   ..

 Align seg 1/1  to: AAY38613  from: 1  to: 318

      33 TTTGCGAACCGCCATGCACATCCTGTTGACCGCCCTGCTCAAATGCCTCT 82
         |||::::::      |||:::|||:::      |||   |||   |||||
      38 PheGlyArgPheMetAlaGluProAlaLeuPheProArgGlnProProLe 54

      83 CCCTGCTGCCGCTTTCCTGTCTGCACACGCTGGGAAACCGGCTCGGACAT 132
         |                        |||:::   :::   :::|||   :
      54 uLeuProAsp.........HisArgHisGlyLysArgThrGlyArgLeuG 68

     133 CTGGCGTTTTACCTTTTAAAGGAAGACCG..................... 161
         ::|||          :::   ||||||
      68 lyGlyGlyArgGlnLysArgLeuArgProTyrValGlyGlyAlaAspAsp 84

     162 ...CGCGCGCATCGTCGCCAA.................TATGCGTCAGG 190
            :::|||||||||||||||
      85 ValHisAlaHisArgArgGlnArgGlnArgMetAlaArgGlnArgProAs 101

     191 CAGGCATGAATCCCGACCCCAAAACGGTCAAAGCCGTTTTTGCGGAAACG 240
          :::::::::||||||:::::   :::   |||   |||
     101 pAlaArgAspGluArgProHisArgArgArgHisArgHisCys....... 115

     241 GCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGA 290
          ||||||      ::::::   ::::::          ::::::
     116 .....ArgArgGlnThrAlaAlaAlaGluIleHisThrAspValAlaPhe 130

     291 CAT.....................................AGAAACAA 301
         |||                                    |||||::
     131 HisAlaCysArgGlnProGlyArgLeuGlnGlnAsnAspCysArgAsnGl 147

     302 TGTTCAAAGCGGTACACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAA 351
         :   :::::::::           |||   :::|||   :::||||||:
     147 nGlnArgGlnAlaTyrAspAlaArgThrPheGlyAlaGluTyrGlyGlnA 164

     352 CACGAAGGGCTGCTATTCATCACGCCGCACATCGGCAG............. 389
         ::         |||   ::::::   :::|||   |||
     164 sn.........AlaProAsnGlnArgThrHisGlyGlnLysProGlnPro 177

     390 ......CTACGATTTGGGCGGACGCTACATCAGCCA.............. 419
                 ::::::||||||      |||||||||
     178 ProArgArgHisIleGlyArgLysProHisGlnProLeuHisAspGlySe 194

     420 ....GCAGCTTCCGTTCCCGCTGACCGC...CATGTACAAACCGCCGAAA 462
             ||||||      |||   :::|||   |||   ::::::|||   :
     194 rHisAlaAlaArgProProGlnAsnArgGlnHisHisArgAlaAlaProA 211

     463 ATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGG 512
         :::::      |||::::::   |||   :::|||   :::||||||||||
     211 spHis......ArgArgGlnAlaAlaIleSerGlnThrGlnArgGlnArg 225

     513 AAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCC 562
         :::      |||            :::         |||:::      ||
     226 AsnProAlaAlaArgProProLeuHisThrAlaProAsnArg......Pr 240

     563 TGCGTTCGGGCGAAGCAACCAT...CGTCCTGCCCGACCACGTCCCCTCC 609
         ||||   :::|||      |||   |||   :::||||||   |||   |
     240 oAlaThrAsnArgArgProHisGlnArgGlnThrArgPro...ProHisP 256

     610 CCTCAAGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGC 659
         ||   |||:::|||   :::         |||      |||:::|||
     256 roHisArgHisArgHisGlnProArgThrGlySerProArgArgThrPro 272

     660 CTATACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAA 709
                     ||||||      ::::::   :::|||      :::::
     273 ProLeuProMetAlaGlyPheProLeuAlaGlnHisGlnTyrAlaSerGl 289

     710 CCCTGTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTG 759
         :               :::   ||||||                     :
     289 yAsnPheArgProArgHisProProAla.....................T 299

     760 CACATCCGCCCCGTCCAAGGGGAATTGAACGGCGACAAAGCCCA 803
         ::||||||||||:::      |||      |||      :::|||
     299 hrHisProProGlnMetAlaGlyCysProArgThrProThrPro 313

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU40508

seq_documentation_block:
ID   AAU40508 standard; Protein; 354 AA.
XX
AC   AAU40508;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #1404.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59512.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 1703; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
```

```
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   354 AA;


alignment_scores:
           Quality:  116.50            Length:    215
             Ratio:  1.165               Gaps:      9
 Percent Similarity:  46.512   Percent Identity: 25.581

alignment_block:
 US-09-303-518D-569 x AAU40508   ..

 Align seg 1/1  to: AAU40508  from: 1  to: 354

     209 CCAAAACGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTTTGGAA 258
         |||   |||   :::|||   |||||||||      :::      ::::: .
      34 ProProArgArgArgProGlnLeuArgLysProGlyArgHisProTyrAr 50

     259 CTTGCCCCCGCGT.................................T 272
         :   |||||||||                                 :
      50 gSerProProArgGlnArgProArgProLeuTrpProAspCysAlaAlaT 67

     273 TTTCAGAAAACCGGAAGACATAGAAACAATGTTCAA......AGCGGTAC 316
         ::|||   :::||| :::|||   ::::::||||||      ||||||
      67 yrSerGlySerArg.AsnHisProArgGlyValGlnGlyTrpSerGlyGl 83

     317 ACGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTA 366
          |||:::               |||:::   :::             :
      83 yArgIleHisArgTrpProArgGlyLeuProArg..............V 95

     367 TTCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAG 416
         ::|||   |||   :::|||:::   |||   |||      |||
      95 alHisProAlaHisGlnArgArgSerArgThrGlySerArgLeu..... 109

     417 CCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCA 466
                                                 ::::::
     110 .....................................AspArgAl 112

     467 AAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAA 516
          :::   |||   :::|||      |||:::   ||||||   |||   :
     112 aThrHisArgAlaHisHisArgGlnGlyArgProSerArgProArgProH 129

     517 ACCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCG 566
         ::|||      :::   :::                |||:::
     129 isArgArgAlaArgProSerGlyTrpLeuGlnProHisArgArgArgPro 145

     567 TTCGGGCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCTCCCCTCAAG 616
            ||||||   :::   :::|||:::::|||           :::||
     146 GlnGlyArgArgGlyGlyHisProThrHisPro...SerAlaAlaGlyAr 161

     617 AAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATACC 666
         |   ||||||||||||   |||:::      |||:::         :::|
     161 gProArgGlyArgArgProGlyAsn......ArgArgLeuHisHisHisH 176

     667 ATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCTGTT 716
         ||   ::::::|||   :::   :::::::::   |||:::
     176 isTrpCysSerGlyHisLeuHisValAsnArgProArgAsn......... 189

     717 TTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACATCC 766
                 |||      :::   |||   ::::::|||      |||||
     190 ......CysArgArgGlnSerHisArgGlyAsnLeuArgGlnHisHisPr 204

     767 G.................CCCCGTCCAAGGGGAATTGAACGG 791

         |                 |||||||||::::::   :::|||
     204 oAlaHisArgArgCysHisProArgProGlnAlaHisHisArg 218

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18664

seq_documentation_block:
ID   AAW18664 standard; Protein; 386 AA.
XX
AC   AAW18664;
XX
DT   24-JUL-1997  (first entry)
XX
DE   Fragmented human NF-H gene +1 frameshift mutant product.
XX
KW   Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
KW   detection; antibody; probe; cancer; neoplasia; neurodegenerative;
KW   Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
KW   Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
KW   PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
KW   cardiovascular; rheumatoid arthritis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1..387
FT                   /note= "X corresponds to a stop codon in the
FT                         accompanying DNA file, AAT69796"
FT   Peptide         129..135
FT                   /note= "antigenic peptide used fro antibody
FT                         production"
XX
PN   WO9712992-A2.
XX
PD   10-APR-1997.
XX
PF   02-OCT-1996;   96WO-IB01106.
XX
PR   11-JAN-1996;   96US-0009832.
PR   02-OCT-1995;   95GB-0020080.
XX
PA   (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
PA   (UYUT-) UNIV STATE UTRECHT.
XX
PI   Burbach JPH,  Grosveld FG,  Van Leeuwen FW;
XX
DR   WPI; 1997-226235/20.
DR   N-PSDB; AAT69796.
XX
PT   Use of mutant genes having frame:shift mutation(s) - for developing
PT   prods. for the diagnosis, prevention and treatment of associated
PT   diseases, e.g. cancer or neuro:degenerative disease
XX
PS   Claim 22; Fig 9; 123pp; English.
XX
CC   AAW18663 and AAW18664 are +2 and +1 frameshift mutations, respectively,
CC   of a sequence comprising fragments of the coding sequence of the
CC   human neurofilament subunit NF-H gene corresponding to nucleotides
CC   1-1162 of the wild-type NF-H gene. This region contains GAGAG motifs.
CC   Frameshift mutants of the tau, ubiquitin, apolipoprotein E,
CC   microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M
CC   and H and amyloid A4 genes are claimed. All these genes share a common
CC   GAGAN motif (N= A, G, C or T), which is the site of common GA
CC   dinucleotide deletion(s) that cause neurodegenerative disorders.
CC   Antigenic peptides used for the production of antibodies, and small
CC   nucleic acid sequences derived from frameshift mutants are used in the
CC   diagnosis, prevention and treatment of cancer and neurodegenerative
CC   diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's
CC   syndrome, frontal lobe dementia (Pick's disease), progressive
CC   supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's
CC   disease, multiple sclerosis, and other degenerative diseases such as
CC   cardiovascular disease and rheumatoid arthritis.
XX
```

```
SQ   Sequence   386 AA;


alignment_scores:
            Quality:  116.50              Length:    327
              Ratio:  0.820                 Gaps:     16
 Percent Similarity:  43.425   Percent Identity: 25.382

alignment_block:
 US-09-303-518D-569 x AAW18664   ..

 Align seg 1/1  to: AAW18664  from: 1  to: 386

      12 ACAATTCAGGCTGTTTCCCCCTTTGCGAACCGCCATGCACATCCTGTTGA 61
         |||   :::::::::::|||   |||               |||:::
      58 ThrAlaLysGlyLeuAspProAlaAlaAlaAlaIleLysGlyProAlaPr 74

      62 C.........CGCCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGT 102
                 |||   |||:::   |||:::||||||:::       |
      74 oTrpSerCysArgSerAlaSerArgProValProAlaSerArgThrCysS 91

     103 CTGCACAC..........................GCTGGGAAACCGGCT 125
         ||:::|||                        ||||||:::|||:::
      91 erGlyHisAspGluLeuArgArgArgGlyArgAlaAlaGlyArgProVal 107

     126 CGGACATCTGGC.........GTTTTACCTTTTAAAGGAAGACCGCGCGC 166
         |||::::::::::              |||::::::::::   |||:::::
     108 ArgAlaAlaAlaTrpArgArgGlnProProLeuArgAlaSerProLysGl 124

     167 GCATCGTCGCCAATATGCGTCAGGCAGG..................... 194
         :   ||||||:::   |||      |||
     124 yTrpArgArgArgAspAlaLeuArgArgTrpLeuLeuGlnArgLeuProL 141

     195 .................CATGAATCCCGACCCCAAAACGGTCAA...... 221
                         :::|||   |||   :::   :::|||
     141 euValAspThrAspValArgGluLeuArgValArgLeuAlaGlnProLeu 157

     222 ........................AGCCGTTTTTGCGGAAACGGCAAAAG 247
                                 |||:::::::|||:::::::::::
     158 ProTrpArgArgArgArgLeuLysHisArgLeuAlaGlyHisAlaGluGl 174

     248 GCGGTTTGGAACTTGCCCCCGCGTTTTTCAGAAAACCGGAAGACATAGAA 297
         :|||   |||              ::::::      |||   |||   :
     174 nArgAlaGlyGly............LeuHisGlyGlyGlyGlyHisLeuT 187

     298 ACAATGTTCAAAGCGGTACACGGCTGGGAACA.................. 329
         :::::   :::   |||:::      |||:::
     187 hrGln***Glu...GlyAlaAlaAlaGlyAlaGluArgProLeuArgArg 202

     330 ..............TGTGCAGCAGGCTTTGGACAAACA........... 353
                       :::|||||||||   :::|||
     203 ValHisArgGlnGlyAlaAlaAlaGlyGlyAlaGlnProGlnProGlyGl 219

     354 .CGAAGGGCTGCTATTCATCACGCCGCACATCGGCAGCTACGATTTGGGC 402
          |||   ::::::         ||||||         ||||||:::||||
     219 yArgGlyCysGlyAlaAlaAlaAlaAlaGlyGlyProLeuArgTyrGlyA 236

     403 GGACGCTACATCAGCCAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAA 452
         ||:::::::::         |||      |||   |||
     236 rgAlaValArg.........AlaArgGlyProArgAspAla......... 246

     453 ACCGCCGAAAATCAAAGCGATAGACAAAATCATGCAGGCGGGCAGGGTTC 502
                            |||::::::   |||   |||:::|||::
     247 .....................ArgArgGlyAlaAlaProGlyArgGlyAl 256

     503 GCGGCAAAGGAAAAACCGCGCCTACCAGCATACAAGGGGTCAAACAAATC 552
         :|||:::            |||
     256 aArgSerAlaThrProGlyAla............................ 263

     553 ATCAAAGCCCTGCGTTCGGGCGAAGCAACCATCGTCCTGCCCGACCACGT 602
            :::||||||   |||:::      :::   ||||||      |||
     264 ..GlyAlaProAlaArgGlyHisArgAlaArgAlaProAlaProArgArg 279

     603 CCCCTCCCCTCAAGAAGGCGGGGAAGGCGTATGGGT.........GGATT 643
               |||:::   |||   ||||||      |||      |||
     280 ArgGlyProAlaAlaArgGlyGlyArgGlyGlyGlyProArgAlaGlyAl 296

     644 TCTTCGGCAAACCTGCCTATACCATGACGCTGGCGGCAAAATTGGCACAC 693
          ||||||                     |||||||||:::   |||:::|
     296 aLeuArg.....................AlaGlyGlyArgGlyGlyAlaA 306

     694 GTCAAAGGCGTGAAAACCCTGTTTTTCTG.........CTGCGAACGCCT 734
         ||         |||:::   :::               :::|||   |||
     306 rgGlyProAlaGluGluGlyAlaGlyAlaAlaGlyGlyValArgLeuPro 322

     735 GCCTGGCGGACAAGGTTTCGATTTGCACATCCGCCCCGTCCAAGGGGAAT 784
         |||                        |||   ||||||      |||
     323 Ala.....................AlaProProProGlyArgGlyGlyAr 332

     785 TGAACGGCGACAAAGCCCATGATGCCGCCGT 815
               |||   :::|||      ||||||
     332 gAlaAlaArgProAspProGlyLeuArgArg 342

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813

seq_documentation_block:
ID   AAB59813 standard; Protein; 1017 AA.
XX
AC   AAB59813;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #4.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1017 AA;


alignment_scores:
            Quality:  115.00              Length:    290
```

```
            Ratio:  0.898                Gaps:     20
 Percent Similarity:  44.138   Percent Identity: 25.517

alignment_block:
 US-09-303-518D-569 x AAB59813   ..

 Align seg 1/1  to: AAB59813  from: 1  to: 1017

      65 CCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTG 114
         |||   ::::::||||||      ||||||   |||:::         ||
      82 ProValThrThrAlaSer......CysArg...ProThrLysProSerTr 95

     115 GGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGC 164
         |:::||||||         ||||||
      95 pLysThrGly......CysTrpArg.......................A 102

     165 GCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAATCCCGACCCCAAAA 214
           :::||||||||||      :::                 |||   |||:::
     102 laSerSerSerProLysSerIleSer............ProLysProArg 114

     215 .........CGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTT.. 253
                  |||   :::|||         :::|||   :::::::::
     115 ProThrCysArgProSerProGlyThrAlaArgArgValSerThrThrSe 131

     254 .......................TGGAACTTGCCCCCGCGTTTTTCAG 278
                                 |||:::   |||   |||   |||
     131 rProArgSerThrThrGlyArgArgTrpSerSerProAlaArgArgSerA 148

     279 AAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTACACGGCTGGGAAC 328
           :::|||                                 ||||||:::
     148 laGlyArg.............................AlaGlyArg 153

     329 ATG...TGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTATTCATCACG 375
             |||:::|||         :::|||:::            |||
     154 AlaGlyCysAlaArgSerSerArgLysThrSerArgProIleArgSerAl 170

     376 CCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCT 425
          |||   |||::::::           :::   ||||||:::|||:::|
     170 aArgProSerCysSer.........LysSerProThrSerValSerAlaP 184

     426 TCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAG 475
         ||            ||||||         |||   ::::      |||
     184 he............ProProSerProAlaArgAlaSerArgThrArg... 195

     476 ACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCT 525
                |||||||||                  ::::::   |||
     196 .........CysArgArg..............AsnSerLeuProSerSe 204

     526 ACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGA 575
         :::   ::::::::::::::::::::   |||      |||   |
     204 rValThrArgSerSerAlaThrArgAlaAlaThrPro......ArgArgL 219

     576 AGCAACCATCGTCCTGCCCGACCACGTCCCCTCCC......CTCAAGAAG 619
         ||   |||           ||||||   ||||||         ::::::
     219 ysThrProCysCysGlyArgThrThrArgProProSerSerThrArgAsn 235

     620 GCGGGGAAGGCG......TATGGGTGGATTTCTTCGGCAAACCTGCCTAT 663
         :::::::::|||           |||   ||||||
     236 SerSerArgAlaThrTrpMetArgTrpAsnSerSer.............. 247

     664 ACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCT 713
               ||||||:::   ::::::   :::                |||:
     248 .......ArgTrpAsnValArgPheProSerMet.........AlaProA 259

     714 GTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACA 763
         ::      |||   :::|||      :::         :::||||||:::
     259 laSerArgAlaProThrAlaLysSerSerArgGlyArgThrIleCysSer 275

     764 TCCGCCCCGTCC..............AAGGGGAATTGAACGGCGACAAA 798
         |||:::||||||               :::::::::      ||||||
     276 SerSerProSerAlaAlaProThrProArgAlaArgThrProAlaThrTh 292

     799 GCCCATG.................................ATGCCGC 812
```

```
            |||                                         ||||
     292 rProThrProSerSerArgGlnProSerGlySerAlaArgProSerProP 309

     813 CGTGTTCAACCGCAATGCCG 832
         ||   |||:::|||:::|||
     309 roSerSerSerAlaIlePro 315

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826

seq_documentation_block:
ID   AAB59826 standard; Protein; 1615 AA.
XX
AC   AAB59826;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #3 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1615 AA;


alignment_scores:
            Quality:  115.00              Length:    290
              Ratio:  0.898                Gaps:     20
 Percent Similarity:  44.138   Percent Identity: 25.517

alignment_block:
 US-09-303-518D-569 x AAB59826   ..

 Align seg 1/1  to: AAB59826  from: 1  to: 1615

      65 CCCTGCTCAAATGCCTCTCCCTGCTGCCGCTTTCCTGTCTGCACACGCTG 114
         |||   ::::::||||||      ||||||   |||:::         ||
     680 ProValThrThrAlaSer......CysArg...ProThrLysProSerTr 693

     115 GGAAACCGGCTCGGACATCTGGCGTTTTACCTTTTAAAGGAAGACCGCGC 164
```

```
693 pLysThrGly......CysTrpArg.....................A 700

165 GCGCATCGTCGCCAATATGCGTCAGGCAGGCATGAATCCCGACCCCAAAA 214
700 laSerSerSerProLysSerIleSer.....ProLysProArg 712

215 .........CGGTCAAAGCCGTTTTTGCGGAAACGGCAAAAGGCGGTT.. 253
713 ProThrCysArgProSerProGlyThrAlaArgArgValSerThrThrSe 729

254 .......................TGGAACTTGCCCCCGCGTTTTTCAG 278
729 rProArgSerThrThrGlyArgArgTrpSerSerProAlaArgArgSerA 746

279 AAAACCGGAAGACATAGAAACAATGTTCAAAGCGGTACACGGCTGGGAAC 328
746 laGlyArg.............................AlaGlyArg 751

329 ATG...TGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTATTCATCACG 375
752 AlaGlyCysAlaArgSerSerArgLysThrSerArgProIleArgSerAl 768

376 CCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGCCAGCAGCT 425
768 aArgProSerCysSer.........LysSerProThrSerValSerAlaP 782

426 TCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAAAGCGATAG 475
782 he............ProProSerProAlaArgAlaSerArgThrArg... 793

476 ACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAACCGCGCCT 525
794 .........CysArgArg..............AsnSerLeuProSerSe 802

526 ACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGTTCGGGCGA 575
802 rValThrArgSerSerAlaThrArgAlaAlaThrPro......ArgArgL 817

576 AGCAACCATCGTCCTGCCCGACCACGTCCCCTCCC......CTCAAGAAG 619
817 ysThrProCysCysGlyArgThrThrArgProProSerSerThrArgAsn 833

620 GCGGGGAAGGCG......TATGGGTGGATTTCTTCGGCAAACCTGCCTAT 663
834 SerSerArgAlaThrTrpMetArgTrpAsnSerSer.............. 845

664 ACCATGACGCTGGCGGCAAAATTGGCACACGTCAAAGGCGTGAAAACCCT 713
846 .......ArgTrpAsnValArgPheProSerMet.........AlaProA 857

714 GTTTTTCTGCTGCGAACGCCTGCCTGGCGGACAAGGTTTCGATTTGCACA 763
857 laSerArgAlaProThrAlaLysSerSerArgGlyArgThrIleCysSer 873

764 TCCGCCCCCGTCC..............AAGGGGAATTGAACGGCGACAAA 798
874 SerSerProSerAlaAlaProThrProArgAlaArgThrProAlaThrTh 890

799 GCCCATG.....................................ATGCCGC 812
890 rProThrProSerSerArgGlnProSerGlySerAlaArgProSerProP 907

813 CGTGTTCAACCGCAATGCCG 832
907 roSerSerSerAlaIlePro 913
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB67173

seq_documentation_block:

ID ABB67173 standard; Protein; 532 AA.
XX
AC ABB67173;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 28311.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL11276.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 28311; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 532 AA;

alignment_scores:

| | | | |
|---|---|---|---|
| Quality: | 113.50 | Length: | 186 |
| Ratio: | 1.158 | Gaps: | 6 |
| Percent Similarity: | 52.688 | Percent Identity: | 26.882 |

alignment_block:
US-09-303-518D-569 x ABB67173 ..

Align seg 1/1 to: ABB67173 from: 1 to: 532

```
183 GCGTCAGGCAGGCATGAATCCCGACCCCAAAACGGTCAAAGC...CGTTT 229
169 AlaSerThrArgGlyArgSerArgArgArgAsnGlyValAlaValSerAl 185

230 TTGCGGAAACGGCAAAAGGCGGTTTGGAACTTGCCCCCGCGTTTTTCAGA 279
185 aSerSerSerGlySerAsnArgValThrSerSerSerAlaAlaGlnGlnA 202

280 AAACCGGAAGA............CATAGAAACAATGTTCAAAGCGGTACA 317
202 rgArgGlyLysSerSerSerSerGluGluGluGluGluGluGluThrSer 218

318 CGGCTGGGAACATGTGCAGCAGGCTTTGGACAAACACGAAGGGCTGCTAT 367
219 AlaGlnAlaThrSerAspAlaSerSerAspGluGluGluArgAlaGlnSe 235
```

```
368 TCATCACGCCGCACATCGGCAGCTACGATTTGGGCGGACGCTACATCAGC 417
    :::   ::::::   :::|||   |||      |||:::   ::::::
235 rAsnAspSerSerProGlnGlnThrArgThrArgArgAlaArgGlnArgG 252

418 CAGCAGCTTCCGTTCCCGCTGACCGCCATGTACAAACCGCCGAAAATCAA 467
          |||         |||   :::::::::   :::|||
252 lnArgLeuSer......AspAspSerAsnIleAsnAspSerGlu..... 264

468 AGCGATAGACAAAATCATGCAGGCGGGCAGGGTTCGCGGCAAAGGAAAAA 517
      |||      |||         |||:::||||||||||:::|||   ::
265 ...AspSerTyrAsnProAsnGluGlyArgGlySerArgArgArgProGl 280

518 CCGCGCCTACCAGCATACAAGGGGTCAAACAAATCATCAAAGCCCTGCGT 567
    :   |||         ::::::   |||||||||   :::|||
280 yAlaAlaArgProSerSerAsnArgGlnThrAsnGlyHisSerSerLysA 297

568 TCGGGCGAAGCAACCATCGTCCTGCCCGACCACGTCCCCTCCCCTCA... 614
        |||  |||      ||||||:::      |||   ||||||
297 rgArgArgLeuAsnLysSerProAlaAsnAlaAlaProGlyProSerLeu 313

615 AGAAGGCGGGGAAGGCGTATGGGTGGATTTCTTCGGCAAACCTGCCTATA 664
      ||||||   ||||||::::::|||         :::   :::
314 ProArgArgThrArgArgIleAlaGlySerSerAlaGluGluAlaAlaSe 330

665 CCATGAC.................GCTGGCGGCAAAATTGGCACACGTC 696
     |||:::                      ::::::||||||   |||:
330 rHisSerGlnHisAspGluSerThrGlnAspSerGlnIleGlyArgArgL 347

697 AAAGGCGT 704
    ::   |||
347 ysGlyArg 349
```